Quaderni di informatica 10

Rassegna di tecnologia ed applicazioni degli elaboratori elettronici - Anno V - Numero 2 - 1978

ARCHIVIO STORICO
ASSOCIAZIONE POZZO DI MIELE
FPDM-53
RQDI-108

Honeywell
Honeywell Information Systems Italia

FPDM - 53

Quaderni di informatica

Rassegna di tecnologia ed applicazioni degli elaboratori elettronici
Anno V - Numero 2 - 1978

Sommario

Le prestazioni degli elaboratori elettronici

La valutazione delle prestazioni di un sistema costituisce un argomento di fondamentale interesse, sia teorico che pratico. Gli autori introducono qui questa tematica, sulla quale hanno scritto un libro di prossima pubblicazione.
D. FERRARI, G. SERAZZI, A. ZEIGNER

Memorie di massa di grande capacità

Le "mass storage facilities", costituite fisicamente da un "alveare" di cartucce di nastro magnetico manovrate automaticamente, rappresentano il sistema di memoria in linea di maggior capacità oggi esistente. L'articolo fa il punto sui problemi inerenti al loro impiego.
D. L. BOYD

Nuove tecnologie di visualizzazione per l'informatica

Il tubo a raggi catodici ha dominato finora il campo dei display per l'informatica. Nuove soluzioni stanno però minando tale predominio, come viene illustrato in questo articolo.
C. DALMASSO

La protezione fisica dei centri di elaborazione dati

Nella società attuale, ogni tipo di attività - dalle industrie ai servizi - è sempre più strettamente dipendente dall'elaboratore. Proteggere queste macchine da danneggiamenti casuali o dolosi risulta quindi un obiettivo di importanza vitale.
R. MIGLIAVACCA

Composizione di musica classica mediante elaboratore elettronico

È questo un impiego originale dell'elaboratore, che trae le mosse dai rapporti tra musica e matematica.
E. GAGLIARDO, P. FORNASARI

Evoluzione e prospettive dell'informatica nelle banche italiane

Le banche costituiscono un settore di punta nell'impiego dell'elaboratore. Qui vengono riassunti i risultati di un'ampia indagine condotta in questo campo.
G. BOSISIO

Direttore Responsabile
F. Filippazzi

Comitato di Redazione
C. Falcetti, P. Lupo, N. Minnaja,
M. Pardi, G. Occhini, G. Rapelli,
F. Ricci, M. Speranza

Redazione
Honeywell Information Systems Italia
Centro di Ricerca e Progettazione
20010 Pregnana Milanese (Milano)

Stampa
La Tipocromo - Milano

Autorizzazione del Tribunale di Milano
n. 93 del 20 Marzo 1974

«Quaderni di Informatica» ospita articoli e contributi di varia provenienza. La responsabilità delle opinioni espresse rimane agli Autori.

La riproduzione di articoli della rivista, o di parte di essi, è consentita solo citando la fonte e l'autore.

Le prestazioni degli elaboratori elettronici (*)

DOMENICO FERRARI
Università di Berkeley

GIUSEPPE SERAZZI
Laboratorio di Analisi Numerica del CNR
Pavia

ALESSANDRO ZEIGNER
Syntax - Milano

1 - Introduzione

Tutti i sistemi dell'ingegneria sono soggetti a valutazioni delle prestazioni in tutte le fasi della loro storia. Durante la progettazione, la costruzione, la vendita, l'esercizio di un sistema, le sue prestazioni vengono valutate da varie persone o enti con finalità e da punti di vista diversi.
In ogni caso, si tratta di verificare se il sistema soddisfa certi requisiti di efficienza e può essere adibito allo scopo per cui lo si vuole utilizzare.
I sistemi informatici non si sottragono, o non dovrebbero sottrarsi, a questa regola. Ovviamente, la verifica del soddisfacimento dei requisiti è tanto più importante quanto più costoso è il sistema che si deve valutare e quanto più critici per la applicazione sono i requisiti di prestazione. Così, le risorse investibili nella valutazione di un singolo microcalcolatore saranno di vari ordini di grandezza più ridotte di quelle che potranno essere investite in uno studio di miglioramento delle prestazioni di una macchina per un grosso centro di calcolo, o nella scelta di un microelaboratore da utilizzarsi in milioni di esemplari.
La valutazione delle prestazioni di un sistema informatico è necessaria non solo durante la vita produttiva del sistema, ma anche durante la progettazione, all'atto della sua scelta da parte di un possibile acquirente o affittuario, e al momento della sua installazione. Si suole infatti classificare le applicazioni delle tecniche di valutazione nelle seguenti tre grandi categorie:
(a) *scelta:* appartengono a questa classe tutti i problemi di valutazione nei quali si deve scegliere un sistema, o alcuni dei suoi componenti, tra un certo numero di alternative esistenti e disponibili (acquisizione di sistemi hardware e software, progettazione di impianti di elaborazione);
(b) *miglioramento:* questa è la classe dei problemi di prestazione che si presentano in sistemi e impianti esistenti e funzionanti; essi saranno discussi diffusamente nel seguito;
(c) *progetto:* in questa classe si raggruppano tutti quei problemi che i progettisti devono affrontare durante la creazione e la realizzazione di un nuovo sistema.
Molte delle tecniche di valutazione sono comuni ai problemi appartenenti a tutte queste classi. Tuttavia nel seguito si farà riferimento soltanto agli studi aventi l'obiettivo (b), che si compiono su sistemi in esercizio, e quindi già progettati, scelti ed istallati e verranno riportate alcune considerazioni sul problema della scelta.
L'obiettivo principale della valutazione delle prestazioni è in questo caso il miglioramento della efficienza del sistema. Tale miglioramento ha di solito effetti difficilmente misurabili ma importantissimi, come per esempio l'aumento dell'indice di gradimento degli utenti, che spesso comporta incrementi di produttività, più rapido sviluppo di nuove applicazioni, e aumenti del numero degli utenti.
Esso può, inoltre, avere conseguenze direttamente e più facilmente misurabili in termini economici, come per esempio le seguenti:

- la riduzione del periodo giornaliero di funzionamento del sistema, con possibile riduzione del numero o dell'impegno degli operatori;
- l'eliminazione o diminuzione dei rinvii di elaborazioni al giorno seguente per saturazione della macchina dovuta a punte di carico;
- una razionalizzazione della configurazione del sistema, che porti ad una configurazione di costo minore (per esempio, la eliminazione di un canale superfluo o di una periferica superflua);
- il rinvio del momento in cui l'aumento del carico saturerà la attuale configurazione e ne imporrà l'espansione o la sostituzione.

Alla luce di queste ultime considerazioni, si può chiarire la precedente affermazione precisando che l'obiettivo della valutazione è in realtà il miglioramento del rapporto costi/prestazioni, cosicchè una diminuzione

(*) Il materiale di questo articolo è stralciato dal volume omonimo che gli Autori hanno in corso di stampa presso l'editore F. Angeli nella "Collana dei Quaderni di Informatica".

dei costi va considerata un risulato positivo per uno studio anche se non è accompagnata da un aumento delle prestazioni. Gli aspetti economici hanno sempre primaria importanza, anche se nel seguito si porrà l'accento soprattutto sugli aspetti tecnici, cioè sulle prestazioni di un sistema.

A fronte dei vantaggi che possono derivare da uno studio di valutazione si deve considerare il suo costo. Prima di intraprendere uno studio, sia i vantaggi che i costi che esso comporterà sono incogniti. Entrambi dovranno essere stimati, anche se ciò può spesso risultare molto difficile. Il massimo vantaggio ottenibile in termini di rapporto costi/prestazioni dipende da quanto lontano è il sistema dalle sue condizioni ottime di funzionamento. Poichè però la nozione di ottimo è in questo caso assai vaga, e poichè tale distanza non è comunque nota (non si conosce infatti generalmente la capacità o potenza di elaborazione massima di un sistema), questa osservazione non è molto utile. Si procede pertanto di solito per tentativi, accontentandosi di stime parziali, che riguardano soltanto l'attività prevista per il futuro prossimo e, pur limitandosi ad essa, lasciano comunque sussistere una quantità non trascurabile di rischio.

Il tipico procedimento con cui si affronta il problema del miglioramento del rapporto costi/prestazioni è, come si vede nella fig. 1, un procedimento iterativo il cui ciclo contiene una fase di *diagnosi* e una di *terapia*. *Lo scopo della fase di diagnosi è quello di stabilire le cause dell'insoddisfacente rapporto costi/prestazioni che caratterizza il funzionamento del sistema. Tali cause sono spesso da identificarsi con le cosiddette strozzature* o *colli di bottiglia (bottlenecks).* Una terapia sarà perciò efficace, in questo caso, se saprà eliminare o "allargare" le strozzature che limitano le prestazioni del sistema. Ovviamente, qualora ci siano più terapie capaci di assicurare il miglioramento desiderato, si dovrà scegliere quella con il più elevato rapporto miglioramento/costo. Poichè alcune terapie possono risultare assai costose (per esempio, quelle che richiedono un'espansione della configurazione, o delicate modifiche al sistema operativo, o cambiamenti del carico), è estremamente utile prevedere gli effetti che queste terapie avranno sul rapporto costi/prestazioni prima che esse vengano applicate.

Alcune delle tecniche utilizzabili nella fase di diagnosi possono, come si vedrà, essere sfruttate anche per la *previsione delle prestazioni.* Si noti che il ciclo diagnosi-terapia del procedimento appena descritto deve talora essere ripetuto, anche più volte, perchè certe strozzature, mascherate da altre più evidenti, possono manifestarsi soltanto quando queste ultime siano state eliminate. Tuttavia, i miglioramenti che via via si ottengono da queste iterazioni risultano in generale decrescenti, e può spesso convenire limitarsi a poche iterazioni. Si noti anche che, con l'uso di tecniche modellistiche di valutazione, è possibile, ma non sempre conveniente, rinviare al termine del procedimento l'effettiva applicazione delle terapie suggerite dalle varie iterazioni.

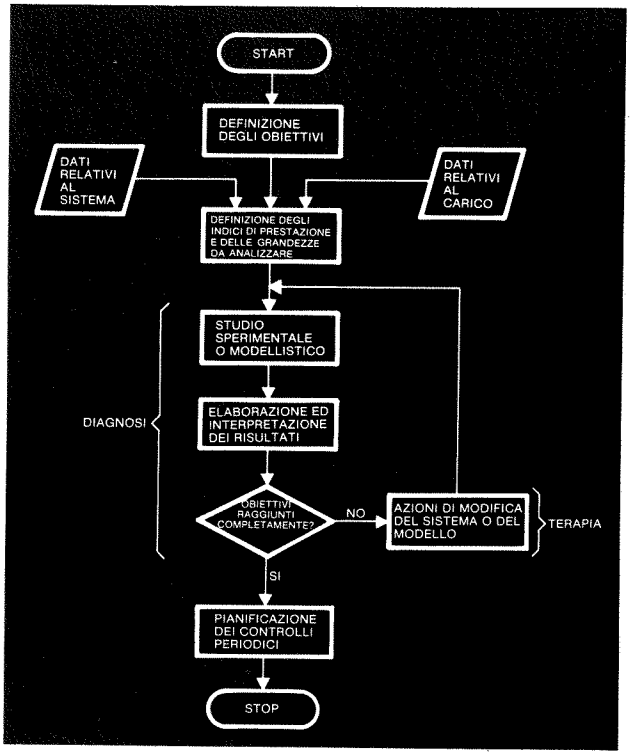


Fig. 1 - Schema di uno studio di valutazione.

2 - La definizione degli obiettivi

Nello schema che rappresenta il ciclo diagnosi-terapia compare, a monte di tutte le altre attività, una fase preliminare di *definizione degli obiettivi.*
La necessità di una definizione degli obiettivi appare evidente non appena ci si appresti a decidere come si debba intraprendere un'analisi delle prestazioni del sistema, a quale grado di approfondimento si debba spingere questa analisi, quante e quali risorse debbano essere investite in essa e, infine, come si possono giustificare tali investimenti. Non è infatti possibile dare una risposta esauriente a queste domande se prima non si cerca di individuare, anche solo approssimativamente, i punti di arrivo di uno studio di valutazione.
Si tratta sostanzialmente di stabilire quali devono essere gli oggetti di indagine e che tipo di informazioni si vuole ottenere riguardo a questi oggetti. Insieme a ciò occorre anche precisare che importanza riveste l'acquisizione delle singole informazioni in modo da poter assegnare delle priorità di indagine e definire la metodologia e gli strumenti di cui fare uso.
Purtroppo nella pratica i motivi che spingono a intraprendere uno studio delle prestazioni nascono frequentemente, più che da una analisi della situazione, da semplici sensazioni o sospetti di malfunzionamento, dalla constatazione che il rendimento del sistema è inferiore alle aspettative oppure dalla protesta degli utenti che lamentano tempi di risposta o di turnaround troppo elevati per le loro esigenze. Disponendo di elementi così generici non si può pensare di affrontare la situazione con il semplice reperimento di uno strumento di misura, a volte assai complesso, e aspettarsi che nel volgere di qualche settimana venga prodotta una documentazione completa di tutte le attività hardware

e software del sistema e suggerita una terapia di sicuro effetto per eliminare i problemi di cui il sistema soffre. Conviene piuttosto porsi qualche obiettivo minore, come per esempio una analisi anche sommaria dei dati raccolti dalle *routines di logging* (attività svolte, tempi impiegati, numero di operazioni I/O eseguite, ecc.), per poter disporre di alcuni elementi iniziali che sono spesso preziosi poichè rendono visibili aspetti e problemi sconosciuti in precedenza. Questi problemi, a loro volta, possono essere classificati in categorie (ad esempio: analisi dell'utilizzo dell'hardware, analisi dell'efficienza dei programmi, analisi del carico, ricerca delle strozzature), che verranno a costituire le aree di indagine in cui si opererà con strumenti e tecniche adeguate.
Al termine di questa analisi preliminare delle caratteristiche del sistema si è normalmente in grado di decidere se gli obiettivi dello studio che si sta per intraprendere devono riguardare aspetti generali della attività oppure aree di applicazione specifiche.
Obiettivi generali sono, ad esempio, i seguenti:
- verificare se esistono le condizioni per evitare, o differire per un tempo determinato, l'acquisizione di nuovo hardware (memoria, unità periferiche o un'altra unità centrale);
- ridurre il carico non produttivo;
- ridurre il carico di lavoro produttivo esistente.

La realizzazione del secondo o del terzo punto permette di aumentare la capacità del sistema, per esempio per fare spazio a nuove applicazioni, quando vi sia già il vincolo di non aumentare l'hardware esistente. Benchè apparentemente simili, essi richiedono due approcci completamente diversi. Ridurre il carico non produttivo significa agire nell'area del sistema operativo e dei programmi di sistema, considerare la riconfigurazione dei componenti hardware e introdurre modifiche operative. Significa in sostanza *ottimizzare il sistema.*
La riduzione del carico produttivo ha senso evidentemetne solo se intesa come miglioramento dei programmi esistenti. È infatti possibile analizzare e ottimizzare i programmi al fine di ridurre il loro impiego di risorse. Tale tecnica di riduzione del carico è particolarmente efficace nei casi in cui la ottimizzazione dei programmi applicativi non ha mai avuto luogo in precedenza. In casi particolari la riduzione del carico produttivo può essere ottenuta spostando parte di esso su un altro sistema.
Il primo, e soprattutto il secondo, degli obiettivi generali sopra indicati richiedono che venga dedicata particolare cura alla *ricerca delle strozzature,* cioè alle cause che limitano la produttività del sistema, e all'individuazione di ciò che può essere fatto per rimuoverle.
Obiettivi che invece riguardano aspetti specifici, e quindi più chiaramente definiti, sono ad esempio:
- abbreviare di una certa frazione il tempo di risposta media ai terminali;
- determinare il livello medio e la durata delle punte del fenomeno di paginazione per verificare la reale necessità di memoria aggiuntiva;
- determinare la migliore ripartizione dei volumi sulle unità disco dei vari canali;
- definire in modo chiaro l'impegno del programma di gestione del *teleprocessing* (TP), in particolare trovare la relazione che lega l'impegno di memoria e di CPU al numero di utenti collegato;
- ottimizzare e bilanciare le attività dei canali.

A volte il raggiungimento di un obiettivo specifico impone la preventiva analisi e risoluzione di altri obiettivi che possono anche non risultare di immediata evidenza. Ad esempio il raggiungimento dell'ultimo degli obiettivi indicati - l'ottimizzazione e il bilanciamento delle attività dei canali - richiederà che si analizzino prima le attività dei singoli canali, determinando anche in che misura queste attività sono sovrapposte tra di loro e con la CPU, poi le attività delle unità periferiche servite dai canali, per determinare in che misura ciascuno di esse contribuisce alla formazione del carico di canale, infine la frequenza degli accessi agli archivi allocati sui volumi montati sulle singole unità ad accesso diretto. Le azioni rivolte al conseguimento dell'obiettivo fissato saranno allora nell'ordine:
a) l'ottimizzazione della disposizione degli archivi sui singoli volumi delle unità ad accesso diretto;
b) la distribuzione ottimale dei volumi sulle diverse unità;
c) l'eventuale riconfigurazione dei collegamenti tra canali e unità periferiche.

L'azione al punto *a)* da sola realizza già un obiettivo specifico, che è quello della minimizzazione dei tempi di accesso agli archivi su disco.
Quelli elencati sono solo alcuni dei molti obiettivi specifici possibili.
Nella scelta dei modi in cui affrontare un dato problema occorre fare attenzione a non dare troppo credito alle soluzioni più ovvie o più intuitive, benchè non di rado la considerazione degli aspetti più appariscenti porti a risultati immediati con un modesto investimento di risorse.

3 - La scelta dello strumento

Stabiliti gli obiettivi dell'indagine che si intende affrontare, è necessario misurare tutto ciò che può eventualemente influenzare questi obiettivi. Misurare, come si è già detto, significa raccogliere dati sulle attività del sistema applicando tecniche adatte e utilizzando opportuni strumenti. Ma di quali tecniche e di quali strumenti è utile e ragionevole servirsi? La domanda ne implica delle altre. Anzitutto, le variabili di cui si è riconosciuto necessario acquisire ulteriori informazioni sono tutte misurabili? E infine, come possono essere misurate in modo sufficientemente accurato? In questo paragrafo vengono svolte alcune considerazioni di carattere generale sulle aree di indagine di uno studio di valutazione delle prestazioni, sui criteri di scelta degli strumenti utilizzabili per eseguire misure in queste aree e sulla misurabilità delle varie grandezze per mezzo di questi strumenti. Dovrebbe così essere possibile, anche sulla base degli argomenti già trattati, dare una risposta sufficientemente esauriente alle precedenti domande.
Le possibili aree di indagine sono:
1) i componenti *hardware* del sistema;
2) il *software,* suddiviso in sistema operativo, programmi di sistema (per la gestione del *teleprocessing,*

per la gestione dei *data base,* di *spool,* di *time sharing,* ecc.) e programmi degli utenti o programmi applicativi;
3) il *carico (workload),* inteso come l'insieme dei programmi che il sistema è chiamato a elaborare in un certo periodo di tempo.
Le attività dei componenti hardware e le attività del software, sia di sistema sia degli utenti, possono essere misurate per mezzo di *strumenti hardware* o di *strumenti software.*
Ciascuno di questi due tipi di strumenti si rivela più adatto dell'altro in determinate aree di misura e, a volte, l'unico possibile per rilevare certi fenomeni.
Qualora si vogliano ottenere informazioni quantitative e descrittive sull'impegno causato dai programmi e non solo dati quantitativi globali sulle attività del sistema, occorre analizzare il carico con strumenti detti, a volte impropriamente, di *accounting.* Si tratta di programmi che elaborano i dati raccolti dalle routines del sistema operativo (funzioni di *logging*) che registrano l'impegno di risorse richiesto da ciascun programma utente. In una certa misura questi strumenti consentono anche di ottenere informazioni su alcune attività dell'hardware e del software di sistema.
La precisione e l'affidabilità delle misure dipende dallo strumento utilizzato per la rilevazione dei dati e, in modo spesso non trascurabile, dal modo in cui lo strumento viene usato e dall'esperienza dell'utilizzatore.
Gli strumenti hardware hanno il pregio di essere indipendenti dal sistema misurato e quindi di non impegnarne le risorse durante le misurazioni. Ciò vale in particolare per la CPU, la cui attività risulta misurabile con estrema precisione da uno strumento di questo tipo poichè esso non risente della non interrompibilità di alcune funzioni del sistema operativo. Essa limita invece la precisione di uno strumento software. D'altra parte, proprio per il fatto di essere indipendenti dal sistema, gli strumenti hardware non sono in grado di fornire informazioni sull'insorgere di condizioni che possono rendere la CPU una risorsa contesa. Infatti

Tabella 1 - Aspetti caratterizzanti gli strumenti hardware e gli strumenti software

	Strumento software campionatore	**Strumento Hardware**
Misure	Quantitative (es.: attività di CPU, attività canali, attività dischi) e descrittive (es.: nome dei programmi in attesa di accedere a un archivio)	Quantitative
Modo operativo	Come programma (utente o di sistema)	Connesso elettricamente al sistema
Impegno CPU	1 - 6% durante il periodo di misurazione	Nessuno
Impegno di memoria	Variabile; da 1 a 30K caratteri	Nessuno
Adattabilità al sistema in esercizio	Deve tenere conto delle caratteristiche dell'hardware e del sistema operativo	Indipendente dall'hardware e dal sistema operativo
Flessibilità a modifiche o ampliamenti nella del sistema	Buona. In genere non è richiesto un adattamento alla nuova configurazione	Limitata dalle caratteristiche dello strumento. Sono necessarie nuove connessioni di sensori al sistema.
Accuratezza delle misure	Dipende dall'ampiezza del campione. Può essere superiore al 99%	100% di accuratezza
Messa in opera	Richiede pochi minuti di uso dell'elaboratore. È subito utilizzabile.	Richiede alcune giornate di lavoro e alcune prove
Addestramento all'uso	Un periodo di 2-3 giorni è sufficiente per apprenderne l'uso e interpretare i risultati	Almeno 2 settimane; prerequisito è una buona conoscenza dell'hardware del sistema
Misure di prestazione del software	È in grado di misurare l'attività dei programmi e dei moduli di sistema; di controllare la presenza delle code e di misurarne la lunghezza	Non rileva l'attività software
Costo	Da 4000 dollari a 25.000 dollari	Da 3 a 10 volte quello degli strumenti software

essi non rilevano il formarsi e il persistere di code di richiesta a tale risorsa.
In generale anzi, questi strumenti non sono adatti a rilevare la presenza di code alle risorse del sistema.
Un'altra caratteristica degli strumenti hardware è la loro flessibilità d'uso. Essa deriva dalla possibilità di combinare in vario modo i segnali provenienti dai sensori. È così consentita la rilevazione contemporanea di un elevato numero di attività concorrenti e quindi la misura del loro grado di sovrapposizione.
Gli strumenti software, che tendono a raccogliere le informazioni in modo seriale, pur non presentando un'analoga flessibilità nella combinazione delle attività dei componenti hardware, sono senz'altro più adattabili degli strumenti hardware ad ampliamenti o modifiche della configurazione del sistema. Essi infatti non richiedono nuovi o diversi collegamenti tra strumento e sistema per rilevare le attività della nuova configurazione, ma solo, eventualmente, qualche lieve modifica ai comandi di estrazione dei dati. L'analista può così disporre di un insieme standard di misure possibili e combinabili tra loro, che gli consente di rendere minimi i tempi di preparazione di una sessione di misurazione.
La tabella 1 riassume le principali caratteristiche dei due tipi di strumenti e la loro applicabilità alla rilevazione delle varie grandezze.

4 - Ricerca delle strozzature (bottlenecks)

Come si è già visto nei paragrafi precedenti, una considerevole parte delle attività tese a migliorare le prestazioni di un sistema riguarda l'individuazione e l'eliminazione delle strozzature. Una *strozzatura* o *collo di bottiglia (bottleneck)* è una limitazione delle prestazioni di un sistema dovuta all'inadeguatezza di un componente hardware o software o dell'organizzazione del sistema stesso. Per brevità, il termine *strozzatura* viene talvolta usato per designare il componente, o la parte del sistema, che provoca una strozzatura. Così, si sente dire che un disco, o un compilatore, sono strozzature, anzichè, come sarebbe più corretto, cause di strozzature.
Una strozzatura si manifesta con un sensibile rallentamento del traffico dei processi in una particolare zona di un sistema. Quando le richieste dei servizi di un certo componente (hardware o software) da parte dei processi costituenti il carico eccedono in intensità e frequenza le capacità di servizio del componente, si creano le condizioni per il verificarsi di una strozzatura. Tuttavia la strozzatura appare realmente solo se gli altri componenti del sistema sono relativamente meno richiesti e lavorano pertanto ad un ritmo meno intenso. Data anzi la natura delle richieste che ciascun processo emette sequenzialmente e non simultaneamente per la grande maggioranza dei componenti, se si crea una strozzatura in una parte del sistema le altre parti risultano assai meno richieste e quindi meno cariche. Infatti molti dei processi attivi finiscono nella coda in attesa del componente sovraccarico e non possono contribuire, mentre sono in attesa, al carico di altri componenti.
In un sistema in cui tutti, o molti, componenti sono sovraccarichi non si possono individuare particolari sorgenti di strozzature: tale sistema, che si dice *saturo*, è evidentemente sottodimensionate rispetto al suo carico. Per migliorare le sue prestazioni deve essere sostituito con un sistema più potente, o si deve trovare il modo di ridurne il carico.
Si parla quindi di strozzature solo quando ad una piccola minoranza (uno o due) dei componenti di un sistema si può attribuire la responsabilità di insufficienti o insoddisfacenti prestazioni. Quando si verifica una strozzatura, qualsiasi provvedimento inteso a migliorare le prestazioni del sistema produce scarsi risultati a meno che non riguardi il componente (o i componenti) all'origine della strozzatura. Si intuisce infatti quale possa essere l'effetto di una CPU più veloce sulle prestazioni di un sistema in cui l'insufficiente velocità di uno dei canali provoca una grave strozzatura. Tale potenziamento, nonostante il suo costo generalmente elevato, non avrà praticamente nessun effetto sulla produttività e sul tempo pesato medio di turnaround del sistema, e contribuirà soltanto a peggiorare la congestione del traffico nel canale. La ricerca delle strozzature è pertanto attività di vitale importanza. Solo agendo sui componenti che creano strozzature si può sperare di ottenere vantaggi (in termini di prestazioni) che giustifichino i costi necessari per il miglioramento di un sistema.

Le precedenti osservazioni ci conducono direttamente ad una estensione della definizione di strozzatura. Quella sopra riportata si applica infatti solo a sistemi multiprogrammati, in cui numerosi processi sono contemporaneamente attivi ed emettono richieste di servizio per i vari componenti del sistema. La definizione estesa, che si potrebbe chiamare *economica* in quanto si basa su un rapporto costi/ricavi, si applica invece anche a sistemi uniprogrammati, a singoli programmi, e più in generale a qualsiasi tipo di sistema(non necessariamente informatico).
Secondo tale definizione, parte di un sistema provoca una strozzatura se il *miglioramento incrementale* delle prestazioni del sistema (opportunatametne definito), per unità di conto investita in quella parte, è massimo.
Sul rapporto costi/ricavi si tornerà al punto 6. Nel seguito di questo articolo per semplicità si discuteranno soltanto sistemi multiprogrammati, e si farà quindi riferimento alla definizione ristretta di strozzatura, che è tuttora quella generalmente adottata nella letteraura tecnica.
Il procedimento concettuale comune ai vari metodi di ricerca delle strozzature è illustrato schematicamente nella fig. 2. Il manifestarsi di sintomi di inefficienza o la periodica verifica delle prestazioni di un sistema conducono il valutatore a chiedersi se vi siano strozzature, dove esse siano, e se sia possibile eliminarle in modo economicamente giustificabile.
Viene quindi formulata un'ipotesi sulla possibile causa della strozzatura. La validità di tale ipotesi viene verificata analizzando i dati già taccolti o raccogliendone degli altri. Quando questi dati confermano un'ipotesi, si presenta il problema di eliminare la causa della strozzatura o almeno ridurne gli effetti. Spesso la ri-

mozione di una strozzatura ne fa apparire un'altra, che può essere studiata e curata secondo lo stesso procedimento, e così via, finchè non si ottiene un sistema *bilanciato*, cioè *privo di strozzature*.
Le strozzature non sono proprietà intrinseche di una certa configurazione hardware-software, ma dipendono fortemente dal carico. Date due configurazioni identiche sollecitate da due carichi diversi, una di esse può avere una strozzatura in un certo componente e l'altra può averla in un componente diverso o non avere nessuna strozzatura. Poichè il carico di molti sistemi varia, anche molto sensibilmente, nel tempo, si possono creare in una data configurazione strozzature transitorie, di durata relativamente breve rispetto a quella di una normale sessione di misura.
Queste strozzature, che si manifestano come temporanee punte di traffico ora in un componente ora in un'altro, possono risultare non facilmente individuabili quando si esaminino i dati cumulativi raccolti dagli strumenti di misura durante un esperimento. Inoltre, correlare questi fenomeni con le loro cause in modo da comprenderli e combatterli efficacemente a distanza di ore o giorni dall'esecuzione delle misurazioni può essere molto difficile. Questi problemi scompaiono o si riducono considerevolmente se il valutatore è in grado di controllare i risultati parziali di un esperimento *in linea*, cioè durante la sessione stessa.
Qualora si sia principalmente interessati a *massimizzare la produttività* di un sistema, i provvedimenti da prendersi in base ai dati di utilizzo devono:
a) *bilanciare il sistema* eliminandone le maggiori strozzature;
b) *massimizzare la sovrapposizione* tra le attività delle varie unità;
c) *minimizzare il tempo di inattività* del sistema.
Questi tre obiettivi sono, come si vedrà nel seguito, distinti ma completamente disgiunti: vi sono infatti modifiche di un sistema che contribuiscono al raggiungimento di più di uno di essi. Se si considera il profilo di utilizzo ci si rende facilmente conto di come i tre obiettivi sopra elencati derivino da quello di massimizzazione della produttività. Dato un certo carico, la produttività con cui esso viene elaborato è infatti inversamente proporzionale alla durata dell'elaborazione, che può essere ridotta sia riducendo il tempo di inattività, sia aumentando la sovrapposizione, sia riducendo il carico delle unità sature (in modo che i tempi di inattività delle altre diminuiscano, e le possibilità di sovrapposizione aumentino).

La tabella 2 contiene alcune ipotesi che si possono formulare nei tre campi definiti dagli obiettivi sopra riportati. Queste ipotesi spesso richiedono, per essere confermate, ulteriori indagini. Talvolta tali indagini si possono condurre con i dati già raccolti mentre altre volte è necessario, o desiderabile, rimisurare il sistema.
In ogni caso, è utile ricordare che, perchè i risultati di una misurazione siano rappresentativi, e quindi l'esperimento sia riproducibile, è necessario che il carico sia *rappresentativo*. Si dovrà pertanto misurare il sistema con un carico pilota che riproduca il più fedelmente possibile il carico reale tipico, o con il carico reale esistente durante periodi di tempo che si ritengono

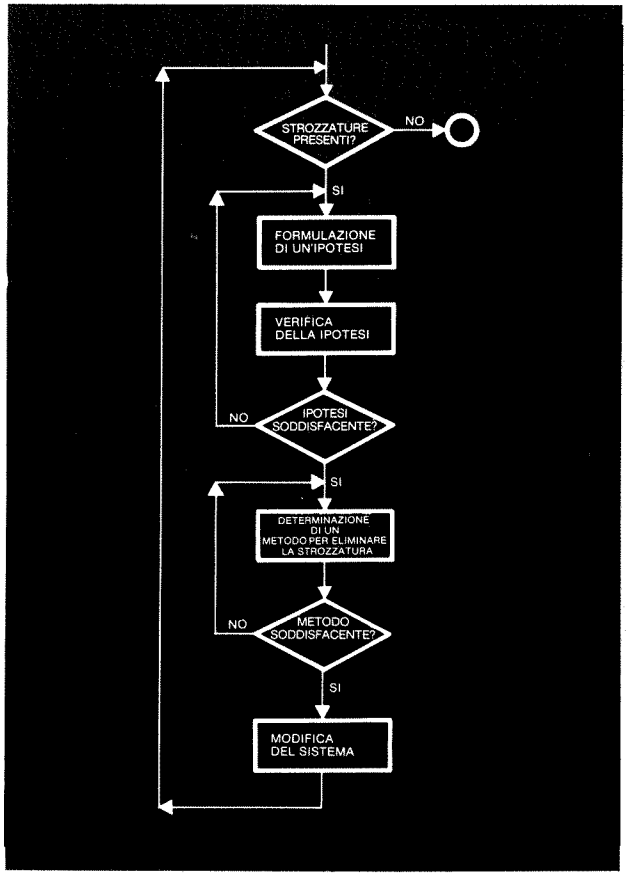


Fig. 2 - Schema per la ricerca ed eliminazione iterativa delle strozzature di un sistema.

tipici. La prima soluzione è potenzialmente meno rappresentativa della seconda ma richiede durate di misurazione assai minori a parità di precisione richiesta. D'altro canto, un carico naturale non esige che il sistema venga dedicato all'esperimento e quindi sottratto ai suoi utenti per la durata delle misurazioni, mentre un carico artificiale è assai meglio controllabile da parte degli sperimentatori.

5 - L'eliminazione delle strozzature

In molti studi di valutazione si cerca di eliminare ogni strozzatura subito dopo averne individuata la possibile causa. Questo accade per esempio quando si usano tecniche di misurazione: in tal caso, infatti, non è possibile scoprire ulteriori strozzature se prima non si è rimossa la principale. Inoltre, quando la ricerca viene effettuata in linea, alcune strozzature (spesso transitorie) possono essere eliminate immediatamente. Non è quindi realistico ritenere che le strozzature e, più in generale, le inefficienze vengano prima scoperte e poi, in un secondo tempo, eliminate tutte insieme. Tuttavia, i tipi di provvedimento che si possono prendere per ottenere questa eliminazione sono concettualmente gli stessi sia che si operi con strumenti di misura *(fuori linea* o *in linea)*, sia che si usino tecniche modellistiche nella ricerca delle strozzature.
Una prima osservazione riguarda la scelta della modifica da apportare al sistema. L'aver confermato l'ipotesi che uno dei dischi del sistema è la causa della strozzatura principale non equivale ad aver definito i parti-

Sintomo	Ipotesi
L'utilizzo di un componente hardware è molto maggiore di quelli di tutti gli altri.	C'è una strozzatura in quel componente. La sua velocità è insufficiente rispetto alla frequenza e alla intensità delle richieste.
La sovrapposizione di alcune delle attività è insufficiente	1) Non ci sono abbastanza lavori che possano essere multiprogrammati. 2) La competizione per certe risorse provoca congestione e quindi una o più strozzature, 3) La schedulazione dei lavori non riesce a mescolare efficacemente più tipi di lavoro.
Il sistema rimane inattivo per troppo tempo.	1) Non ci sono abbastanza lavori; 2) Troppo tempo viene perduto nel montare e smontare dischi e nastri; 3) Il tempo di ricerca non sovrapposto ad altre attività è eccessivo.

Tabella 2 - Alcune ipotesi che i dati di utilizzo possono suggerire.

colari provvedimenti che si debbono adottare. Il sistema in questo caso può essere ribilanciato in vari modi, tra i quali i seguenti:
1) sostituzione dell'unità a dischi con un'unità più veloce;
2) spostamento di alcuni degli archivi più richiesti ad altri dischi da montarsi su unità meno cariche;
3) riorganizzazione delle allocazioni sul disco in modo da ridurre i movimenti del braccio;
4) modifica della politica di schedulazione dell'unità in modo da ridurre i movimenti del braccio, o le attese dovute alla latenza del disco, o entrambe tali cause di inefficienza.
Si noti che questi provvedimenti non sono mutuamente esclusivi. Al contrario, la strozzatura causata dal disco può essere curata mediante l'applicazione simultanea di più terapie. I costi e i rischi di queste cure variano della cura. È perciò opportuno scegliere bene i provvedimenti da adottare e, ancor prima di far ciò, stabilire se sia conveniente o no intraprendere un'azione per eliminare o ridurre la strozzatura. Queste decisioni dovranno essere prese basandosi sui risultati di ulteriori indagini (per esempio, misurazioni supplementari).
Infatti, l'aver scoperto che una strozzatura è dovuta ad una delle unità a dischi non è affatto sufficiente ai fini delle decisioni da prendersi per eliminarla.
Vi sono due classi principali di modifiche che si possono apportare ad un sistema per eliminarne una strozzatura:
a) modifiche alla configurazione hardware, che consistono nell'aggiunta, sostituzione o talora anche eliminazione di uno o più componenti hardware;
b) modifiche che non si ripercuotono sulla composizione del sistema ma ne cambiano in qualche modo l'organizzazione.
Le modifiche appartenenti a queste due classi possono essere chiamate, generalizzando termini che sono spesso usati con con significati più ristretti, provvedimenti di *upgrading* (o potenziamento) e di *tuning* (o accordatura), rispettivamente. I primi sono generalmente più visibili e più radicali, i secondi possono essere più rischiosi se richiedono modifiche del software di sistema non previste dai suoi progettisti. Il costo è spesso, ma non sempre, minore per i provvedimenti di tuning che per quelli di upgrading. Le modifiche più costose, o più rischiose, dovrebbero comunque essere attuate soltanto quando i valutatori sono ben sicuri dell'entità e della convenienza degli effetti provocati da tali modifiche. In altre parole si dovrebbe poter prevedere con buona affidabilità la loro influenza sulle prestazioni del sistema. Questo è il tipico problema che un modello del sistema consente di risolvere in modo completo e preventivo (ovviamente perchè la soluzione sia soddisfacente, il modello deve essere affidabile). si può, per esempio, ricorrere ad un simulatore commerciale o ad un simulatore costruito proprio a questo scopo dai valutatori o da un modello analitico. Alternativamente, si può assumere il rischio e modificare il sistema senza aver condotto nessun studio previsionale degli effetti di tali cambiamenti. Nel far ciò, un esperto valutatore *(performance analyst)* sarà di solito confortato dalla sua conoscenza del sistema, del carico e dei vantaggi che le principali azioni di tuning sono in grado di offrire.

6 - Costi e ricavi di uno studio di valutazione

I benefici che uno studio di valutazione delle prestazioni può arrecare ad un impianto informatico e all'organizzazione nella quale l'impianto opera non sono generalmente gratuiti. La pianificazione e l'esecuzione di uno studio comportano sempre un costo che dovrebbe essere stimato e confrontato al possibile ricavo prima di intraprendere lo studio medesimo. In tale confronto dovrebbero naturalmente essere considerati anche i costi associati alle azioni di modifica del sistema che verranno suggerite dallo studio. Mentre alcune delle azioni possibili sono relativamente poco costose (per esempio, il riordinamento degli archivi memorizzati su un disco, la rigenerazione del sistema operativo con parametri diversi, lo spostamento di un'unità di memoria secondaria da un canale ad un'altro, e così via), altre possono esserlo molto di più (per esempio, il raddoppio della capacità della memoria primaria, il potenziamento della CPU, l'aggiunta di canali, unità di controllo, e così via). Prevedere questi costi prima di aver compiuto lo studio è molto difficile; ancor più arduo è prevedere gli effetti causati sul sistema dalle azioni che uno studio suggerisce.
Ovviamente nello stimare i benefici che derivano da tali effetti al centro elaborazione dati e all'organizza-

zione di cui esso fa parte non si può quasi mai procedere oltre il livello delle congetture.
Tuttavia, le inevitabili difficoltà di una valutazione preventiva dei costi e dei ricavi non devono nè possono giustificare una rinuncia a compiere, sia pure in modo approssimativo e incompleto, tale valutazione.
Non bisogna dimenticare che uno studio di valutazione è un investimento e come tale va considerato.
I metodi di stima e di controllo della bontà degli investimenti sono tra le fondamentali tecniche moderne di gestione. Poichè ad ogni investimento è di solito associato un rischio, si preferisce sempre, e con ragione, un rischio calcolato ad un rischio incognito.
Quali sono dunque le principali componenti del costo di uno studio di valutazione? Come si vede nella tabella 3, esse possono essere classificate nelle seguenti categorie:
(a) *costi del personale* che dovrà compiere lo studio e di quello che dovrà eseguire le modifiche suggerite da esso;
(b) *costi di acquisizione e manutenzione* degli strumenti di misura da usarsi nello studio;
(c) *costi di esercizio* della strumentazione per la raccolta e la riduzione dei dati;
(d) *costi delle modifiche* che si devono apportare al sistema per migliorarne le prestazioni (o meglio il rapporto costi/prestazioni).
La tabella 3 contiene anche alcune delle più importanti 'voci di costo' che possono comparire in ciascuna di queste categorie.
I costi nella categoria (b) e (c) sono i più facilmente prevedibili, mentre quelli della categoria (d) sono, come si è detto, i più difficili a stimarsi con una ragionevole precisione.
I possibili benefici di uno studio di valutazione sono, come si è detto, di stima ancora più incerta e disagevole. Alcuni di tali benefici, come per esempio la maggior 'soddisfazione' degli utenti che di solito fa seguito ad un miglioramento dei tempi di turnaround o di risposta, sono intangibili o, almeno, molto difficilmente traducibili in vantaggi di carattere economico per l'organizzazione; è vero che un aumento della 'soddisfazione' provoca spesso un aumento della produttività dei programmatori, ma anche questo è difficilmente misurabile ed ancor più difficilmente prevedibile. In alcuni casi tuttavia, i benefici sono almeno parzialmente quantificabili con grande facilità; si pensi per esempio al caso di uno studio che mostri come una o più unità hardware possano essere eliminate senza pregiudicare le prestazioni del sistema.
Come un certo miglioramento di *efficienza* di un sistema informatico si tramuti in un miglioramento di *efficacia* (cioè dei benefici apportati dall'informatica ad una organizzazione) dipende dal tipo di organizzazione. Per esempio, la velocità di fatturazione di certi beni o servizi può avere un notevole peso sui profitti di una azienda che produce o distribuisce tali beni o servizi, specialmente quando il costo del denaro è elevato; ancora, la prontezza con cui la situazione di cassa viene portata a conoscenza della direzione di un'impresa può avere importanti effetti sul benessere finanziario dell'impresa stessa. Poichè i benefici per un'intera organizzazione sono così diversi, variabili nel tempo e difficilmente classificabili, ci si limiterà a discutere i principali vantaggi che un centro elaborazione dati può attendersi da uno studio di valutazione.
Alcuni di tali vantaggi sono elencati nella Tabella 4.
Occorre osservare a questo punto che il raggiungimento di sensibili vantaggi di questi o di altri tipi non è affatto garantito.
Quando un sitema funziona con prestazioni vicine a quello ottime corrispondenti al carico cui esso è sottoposto, l'unico modo di migliorarne le prestazioni è quello di migliorarne il carico. Quando anche questo sia stato ottimizzato, qualsiasi studio di valutazione diventa non solo inutile ma dannoso, poichè assorbe risorse senza produrre alcun risultato. È necessario però aggiungere che, nonostante quanto si è appena detto, gli studi di valutazione sono nella maggior parte dei casi utili principalmente per le seguenti ragioni:
(1) molti degli impianti informatici funzionano con prestazioni lontane da quelle ottime;
(2) un impianto ottimizzato rimane in condizioni vicine all'ottimo per un tempo abbastanza lungo solo quando il suo carico è molto stabile; nella maggior parte degli impianti il carico non è stabile ma generalmente cresce e talvolta anche le sue caratteristiche si

Categorie di costo	Esempi di voci di costo
Personale	Valutatori Programmatori di sistema Programmatori applicativi Addestramento
Acquisizione e manutenzione strumenti	Acquisto o affitto di: - strumenti hardware - strumenti software - strumenti di accounting Simulatori Costi di installazione Contratti di manutenzione
Raccolta e riduzione dei dati	Consumi di CPU Consumi di memoria Consumi di I/O Costo dello spazio di memoria secondaria Costo del sistema (nel caso di esperimenti con sistema tutto dedicato)
Azioni sul sistema	Potenziamento (CPU, canali) Ampliamenti (memorie) Rigenerazione del sistema Riconfigurazione Riordino di archivi Verifica della correttezza dei programmi modificati Ristrutturazione dei programmi

Tabella 3 - Principali componenti del costo di uno studio di valutazione

Riduzione dei costi di personale e di affitto del sistema dovute alla diminuzione del periodo di funzionamento giornaliero
Rinvio del potenziamento e ampliamento del sistema
Riduzione dei costi d'affitto dovuto all'eliminazione di componenti poco utilizzati
Riduzione del volume di elaborazioni svolte in altri centri di calcolo per smaltire eccedenze di carico
Aumento della produttività dei programmatori dovuto al miglioramento dei tempi di risposta o di turnaround
Possibilità di aumento del carico con incremento dei profitti per il centro di elaborazione o riduzione dei costi per ciascun utente
Possibilità di assorbimento di più alte punte di carico da parte del sistema a parità di degradazione della qualità del servizio.

Tabella 4 - Alcuni dei vantaggi che possono derivare da uno studio di valutazione

modificano nel tempo;questo spiega perchè l'affermazione (1) sia vera anche per molti degli impianti per cui si siano già compiuti in passato studi di valutazione.
Tutte le tecniche note per la valutazione degli investimenti possono essere applicate al problema di stabilire se sia conveniente iniziare uno studio delle prestazioni di un sistema. Queste tecniche permettono di confrontare i benefici che ci si aspettano da attività di valutazione con quelli che si possono ottenere investendo le risorse in altri settori dell'attività dell'organizzazione o del centro elaborazione dati, e di scegliere gli investimenti che, a priori, appaiono più redditizi. I più semplici di tali metodi sono quelli basati sul *tempo di ripagamento* o *di ritorno dell'investimento* e sul *tasso interno di ritorno.*
Il tempo di ripagamento è definito come il tempo necessario a far sì che i ricavi pareggino i costi *(break-even point);* raggiunto tale tempo, i ricavi (al netto delle spese di esercizio e di manutenzione, se ve ne sono) si tradurranno direttamente in profitti. La curva che descrive l'andamento nel tempo dei profitti (o delle perdite) dovuti ad un investimento in attività di valutazione delle prestazioni assume spesso una forma come quella del diagramma a tratto continuo in fig. 3 (si noti la differenza rispetto alla curva a tratti, che rappresenta l'andamento di un tipico progetto applicativo nel campo informatico).

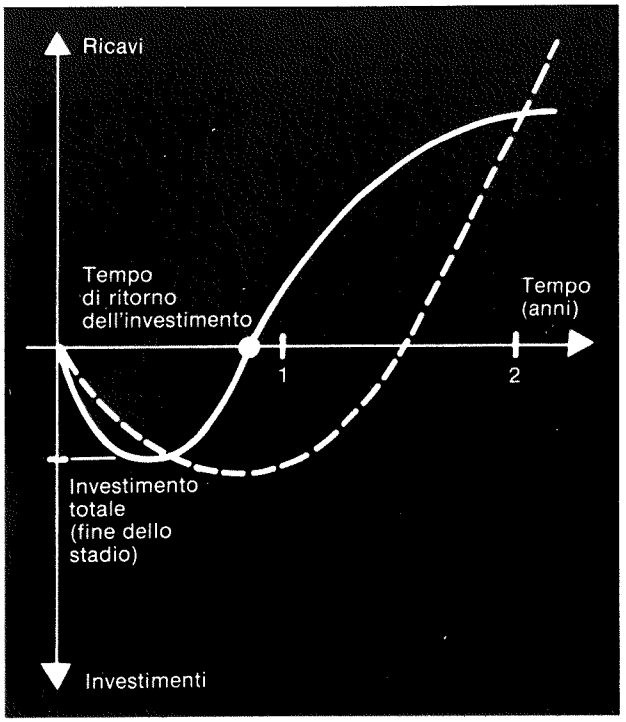


Fig. 3 - Andamento nel tempo dei ricavi in uno studio di valutazione (linea continua) e in un progetto applicativo (linea a tratti).

Se non si tiene conto del costo del denaro, che può influenzare il risultato poichè gli investimenti vengono fatti prima e in breve tempo mentre i ricavi si manifestano più tardi e con maggiore lentezza, il tempo di ritorno dell'investimento t_{rit}, in anni, può essere calcolato con la formula:

$$t_{rit} = \frac{i}{r_a} \qquad (8.1)$$

dove i è l'investimento totale e r_a il ricavo annuo.
Vi sono vari modi di calcolare il tasso interno di ritorno. Una semplice definizione è quella che lo identifica con il ricavo per unità di tempo dovuto ad un investimento. Secondo questa definizione, che ignora anch'essa (a meno che non se ne tenga esplicitamente conto) gli effetti finanziari dei diversi tempi delle entrate e delle uscite, il tasso interno di ritorno è la derivata della curva in figura 3. Questa derivata è, in generale, una funzione del tempo.

7 - Bibliografia

Una bibliografia esauriente per chi voglia approfondire i singoli temi di questo articolo è riportata nel volume sull'argomento in corso di pubblicazione nella "Collana dei Quaderni di Informatica", presso l'editore F. Angeli. Si rinvia pertanto a tale fonte il lettore interessato.

Memorie di massa di grande capacità

D. L. BOYD
Honeywell Corporate Technology Center
Minneapolis (USA)

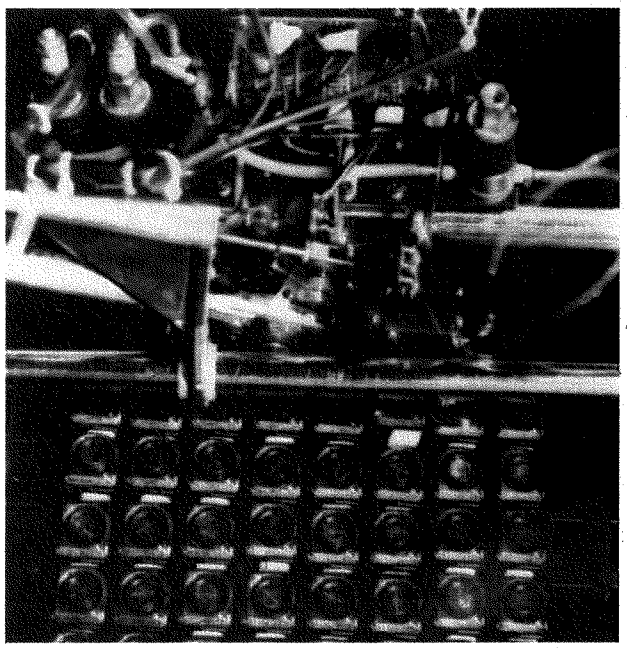
1 - Introduzione

La comparsa sul mercato di dispositivi di memoria ausiliaria in linea di grandissima capacità - le cosidette "Mass Storage Facilities" (MSF) - ha posto nuovi e interessanti problemi. Questi problemi includono nuove applicazioni dei dispositivi per la memorizzazione dei dati in linea, la necessità di nuove strutture di dati organizzate intorno alle caratteristiche fisiche delle MSF, problemi di efficienza, il miglioramento del controllo automatico dei dispositivi fisici e del software di gestione delle memorie, un'aumentata preoccupazione per l'integrità dei dati, e la necessità di maggiore protezione della memoria.
Tenendo conto di questi problemi, sono state considerate diverse soluzioni per incorporare una memoria MSF nel sistema ospite. Questi approcci hanno reso evidente l'esigenza di maggiori ricerche, in particolare nell'area della gestione della memoria e delle strategie di scheduling. La tecnologia della memoria si sta infatti sviluppando rapidamente, mentre le realizzazioni pratiche sono ostacolate da molti problemi.
L'articolo delinea tre di questi approcci dal punto di vista dei sistemi automatici di gestione della memoria ausiliaria. (AMMS: Auxiliary Memory Management System). Alla fine, viene proposto un semplice modello di gestione e presentati alcuni problemi di ricerca ancora aperti.

2 - Caratteristiche delle memorie di massa di grande capacità

Le memorie MSF sono costituite da un grandissimo numero di cartucce di nastro magnetico che risiedono in una libreria con struttura a celle (fig. 1). La capacità di memorizzazione delle cartucce va da 8 milioni a 50 milioni di byte, mentre la libreria di un MSF contiene da 2000 a 4720 cartucce. Un sistema può avere una capacità di memoria MSF che arriva a quasi 500 miliardi di byte. Sotto il controllo del programma, qualsiasi cartuccia può essere rimossa dalla libreria e spostata ad una stazione di lettura/scrittura, dove viene montata e resa disponibile per l'uso.
Questa operazione di caricamento richiede da 3 a 13 secondi in dipendenza del dispositivo, della distanza della cella della cartuccia indirizzata dalla stazione di lettura/scrittura, e a seconda la posizione del meccanismo di accesso all'inizio del ciclo. Una volta che la cartuccia è montata, il trasferimento dei dati è molto rapido, paragonabile a quello del disco o del tamburo. Tuttavia, la registrazione di dati su questo mezzo di memoria è paragonabile alla registrazione su un nastro magnetico o su altri dispositivi ad eccesso sequenziale, e può richiedere quindi il lungo tempo di una ricerca sequenziale, cioè molti secondi, o soltanto il tempo di avvio di un nastro, cioè alcuni millisecondi.

Fig. 1 - Particolare di una Mass Storage Facility (38510), che mostra una parte della libreria di cartucce ed uno dei meccanismi di prelievo delle medesime.



Relativamente poche stazioni di lettura/scrittura sono impiegate per realizzare l'accesso a un numero molto grande di cartucce. Ciò pone dei limiti a richieste concomitanti di accesso ai dati.
Inoltre, essendo il mezzo di registrazione delle MSF un nastro magnetico, si hanno limitazioni nel numero di operazioni di lettura/scrittura effettuabili con una stessa cartuccia. Il basso costo di quest'ultima compensa in parte questa limitazione. Le cartucce consumate o danneggiate possono essere sostituite; ma se tale onere non è tollerabile occorre mantenere basso il numero di operazioni di lettura/scrittura su ogni cartuccia.
Riassumendo, le memorie di massa MSF sono caratterizzate da: capacità di memorizzazione molto ampia a relativamente basso costo, capacità di controllo automatico, tempo di montaggio di una cartuccia relativamente breve, tempo di ricerca paragonabile a quello dei dispositivi ad accesso sequenziale, alta velocità di trasferimento dei dati, piccolo numero di meccanismi di lettura/scrittura per accedere a grandi raccolte di dati, e infine limitazioni nel numero di operazioni di lettura/scrittura effettuabili su uno stesso substrato di memoria.

3 - Applicazioni delle memorie di massa di grande capacità

Il costo per bit di una MSF è molto basso per le applicazioni i cui requisiti di memoria giustifichino il dispositivo; infatti, la capacità di memorizzazione di una MSF deve essere altamente utilizzata per raggiungere un conveniente livello di costo.
Una applicazione ovvia delle memorie MSF è la sostituzione di unità a nastro magnetico che registrano ampi flussi sequenziali o convertono flussi ad accesso random in memoria d'archivio. I vantaggi sono un montaggio e smontaggio veloce e automatico del mezzo di memorizzazione, il controllo automatico del flusso usando i cataloghi del sistema, un'altissima densità di impaccamento con trasferimento dei dati ad alta velocità. La capacità e il costo delle cartucce si possono ritenere paragonabili a quelli del nastro magnetico. Infatti, raramente un'intera bobina di nastro viene usata per memorizzare un solo flusso e raramente c'è più di un flusso su un nastro. Così di solito il nastro non viene usato interamente, mentre la cartuccia sarà sfruttata di più. Però, il costo del sistema e del dispositivo MSF è molto più alto di quello delle unità a nastro. Chiaramente i vantaggi devono essere attentamente soppesati rispetto a questo costo quando si considera l'MSF come dispositivo di memoria d'archivio.
Lo svantaggio maggiore è nel limitato numero di stazioni di lettura/scrittura associate con ciascun MSF. Questo limita pesantemente il numero di flussi a cui si può accedere contemporaneamente.
Un altro tipo di applicazione delle memorie MSF è la memorizzazione ed il reperimento di flussi altamente strutturati, come il flusso partizionato, le cui partizioni sono sottoarchivi o blocchi di dati. Ciò a causa delle limitate capacità di accesso random delle MSF. Le strutture fisiche dei dati di alcuni tipi di flussi possono essere migliorate dall'organizzazione della registrazione imposta dalla cartuccia di dati. Ciò permette l'accesso diretto a certe posizioni di un flusso senza bisogno di accedere all'intero archivio.
Un terzo tipo di applicazione è un sistema di memoria gerarchica. Le caratteristiche delle MSF le pongono naturalmente al livello più basso nella gerarchia delle memorie, al livello cioè sottostante le unità a dischi magnetici. Quando l'informazione memorizzata a questo livello viene richiesta per l'elaborazione, viene spostata seguendo la gerarchia con un processo chiamato "staging". Finchè l'attività dell'informazione rimane alta, l'informazione scende dai livelli più alti o viene ricopiata ai livelli più bassi da un processo chiamato "destaging" o "archiving". Questa applicazione delle MSF è interessante, poichè si adatta perfettamente all'impiego in sistemi gerarchici di memoria. Inoltre, ciò permette una memoria logica ad un solo livello con gestione automatica della memoria per tutti i livelli di memoria primaria o ausiliare. Sebbene questa sia un'applicazione interessante delle MSF, l'approccio realizzato deve essere molto cauto per evitare problemi molto seri di prestazioni. Come con i sistemi con gestione automatica della memoria, il livello di prestazioni del sistema dipende infatti molto dalle strategie scelte per l'implementazione.
Il tempo di accesso iniziale relativamente alto e il limitato numero di stazioni di lettura/scrittura impediscono l'impiego di MSF in tempo reale o in ambiente interattivo, dove i dati devono essere disponibili immediatamente. Tali applicazioni richiedono che l'informazione sia predisposta su un dispositivo che richiede un tempo d'accesso minore, come un disco, un tamburo o la memoria principale. L'MSF può essere collegata a un sistema in tempo reale ma non può essere usata in tempo reale.

4 - Soluzioni realizzative

Ci sono tre modi per incorporare il controllo di una MSF in un sistema operativo ospite: distribuito, gerarchico od una combinazione di questi.
Tutti e tre gli approcci definiscono un sistema di gestione della memoria ausiliare (AMMS, Auxiliary Memory Management System), che è responsabile della gestione delle informazioni per tutta la memoria ausiliare. Ma anche con l'AMMS, certe linee di gestione possono influenzare grandemente le prestazioni del sistema. I dettagli della strategia migliore dipendono dalla struttura e dalle applicazioni delle informazioni memorizzate nella memoria ausiliare. È necessaria una maggiore informazione sotto forma di modelli, risultati e tecniche di analisi.

a) Soluzione distribuita
L'approccio distribuito mette l'MSF allo stesso livello degli altri dispositivi di memoria ausiliare (Fig. 2); le interfacce di input/output standard trasmettono le informazioni direttamente tra l'MSF e la memoria principale. Questa configurazione è adatta ad applicazioni

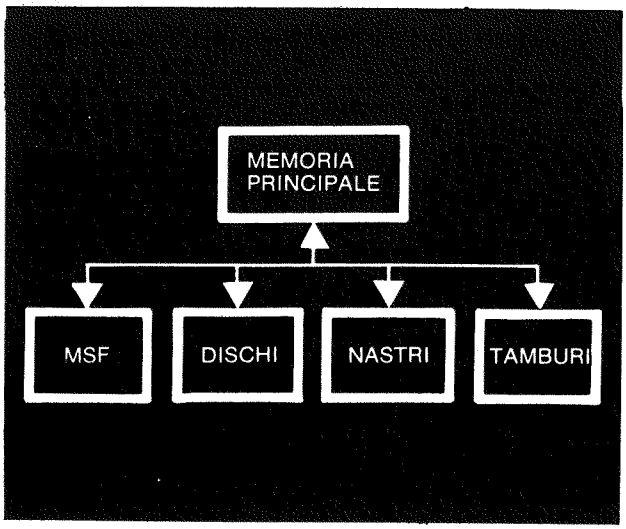


Fig. 2 - La MSF come memoria ausiliaria di primo livello logico.

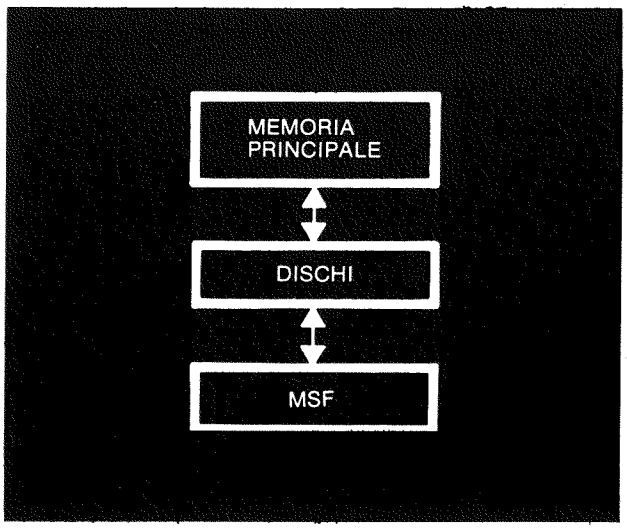


Fig. 3 - La MSF come memoria ausiliaria di secondo livello logico.

che altrimenti potrebbero usare il nastro magnetico, cioè applicazioni che richiedono solo l'accesso sequenziale alla informazione memorizzata. Un esempio è dato da un sistema di documentazione che reperisce le voci del testo in modo sequenziale. Inoltre alcuni archivi strutturati in modo particolare possono essere memorizzati nel MSF per accesso diretto, come è stato descritto nel paragrafo precedente. L'approccio distribuito è altamente interessante per alcune applicazioni, perchè esso associa la portabilità, la sicurezza e il basso costo dei dispositivi a nastro. Le prestazioni sono migliori di quelle dei sistemi convenzionali a nastro magnetico perchè gli archivi vengono ripristinati automaticamente e le cartucce vengono montate e smontate automaticamente. Inoltre, se l'elaborazione dell'informazione memorizzata è sequenziale, i meccanismi di start/stop danno certamente un ritardo minore della latenza rotazionale e del tempo di ricerca della traccia delle unità a disco. Flussi sequenziali che sono stati memorizzati su disco possono essere posti su MSF, rendendo il disco libero per maggior capacità di memoria ad accesso random. Bisogna comunque essere consapevoli del tempo di accesso iniziale relativamente alto e del fatto che la possibile alta competizione per l'uso dei meccanismi di lettura/scrittura possono impedire una tale scelta.

L'AMMS per configurazioni MSF distribuite ha un'interfaccia con il sistema convenzionale di flusso del sistema operativo ospite. Il flusso e le sue voci sono indirizzati per nome permettendo all'AMMS di nascondere le caratteristiche del supporto fisico attraverso le interfacce del sistema di flusso ospite. Inoltre, la strategia di gestione primaria dell'AMMS è una strategia di allocazione; essa può gestire automaticamente la memoria ausiliare scegliendo il dispositivo di memoria ausiliare appropriato. Si richiede un algoritmo per scegliere tra l'MSF e le altre forme di memoria ausiliare per l'allocazione dei flussi. I fattori che determinano la strategia di allocazione comprendono la frequenza d'uso del flusso, la tecnica di accesso, la dimensione del flusso e il numero di flussi che richiedono accesso contemporaneamente in qualunque momento. Un eccessivo pre-svuotamento della stazione di lettura/scrittura, o un uso eccessivo di flussi molto piccoli può provocare un alto volume di traffico delle cartucce, il che diminuisce l'efficienza del sistema.

L'algoritmo di allocazione del flusso dipende dall'applicazione. Se progettato non correttamente, l'algoritmo può provocare severe variazioni delle prestazioni del sistema.

b) Soluzione gerarchica

Il secondo approccio realizzativo tratta l'MSF come un dispositivo di memoria ausiliaria di secondo livello in una gerarchia di memoria (Fig. 3). L'AMMS suppone che un disco sia posizionato al livello medio, e che la memoria principale e le altre memorie ad alta velocità siano al livello più alto. Tuttavia le forme correnti di AMMS trattano l'MSF e il disco come un unico livello logico di memoria ausiliaria, in cui l'MSF è in sostegno al disco. Questo è lo stesso modo in cui gli attuali sistemi di memoria virtuale realizzano una memoria di lavoro a un livello mediante una gerarchia a due livelli, con la memoria ausiliaria come memoria di sostegno alla memoria principale. Questo approccio permette un'interfaccia conveniente tra la memoria virtuale ausiliaria e i sistemi di flusso convenzionali. In questo modo la gestione automatica della memoria può essere estesa all'intera gerarchia dei dispositivi di memoria.

La gestione di una memoria ausiliaria virtuale, come quella dei sistemi a memoria virtuale, comprende tre strategie, di caricamento, allocazione e riallocazione, Sono state tentate parecchie variazioni di queste strategie. Sebbene sia stata condotta una gran quantità di ricerche sulle strategie per i sistemi gerarchici di memoria principale, esse sono soggette ad alcune variazioni interessanti nei sistemi gerarchici di memoria ausiliaria.

c) Soluzione mista distribuita e gerarchica.

In un approccio misto (Fig. 4), l'MSF è direttamente accessibile dalla memoria principale per alcune applicazioni, pur rimanendo al livello più basso della gerarchia (al di sotto del disco) per altre.

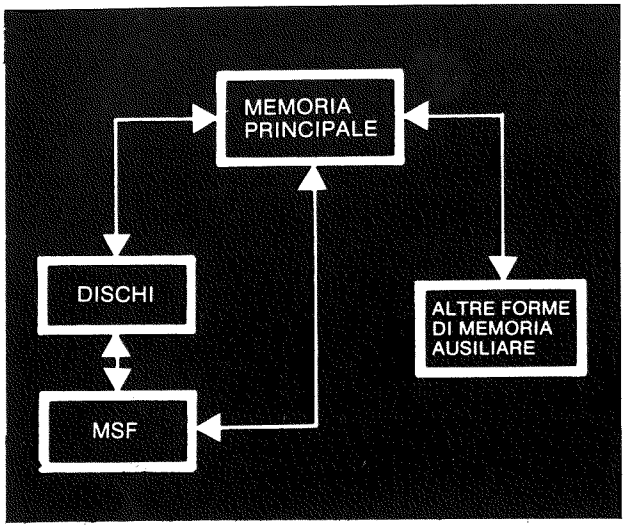


Fig. 4 - La MSF come memoria ausiliaria mista, di primo e secondo livello.

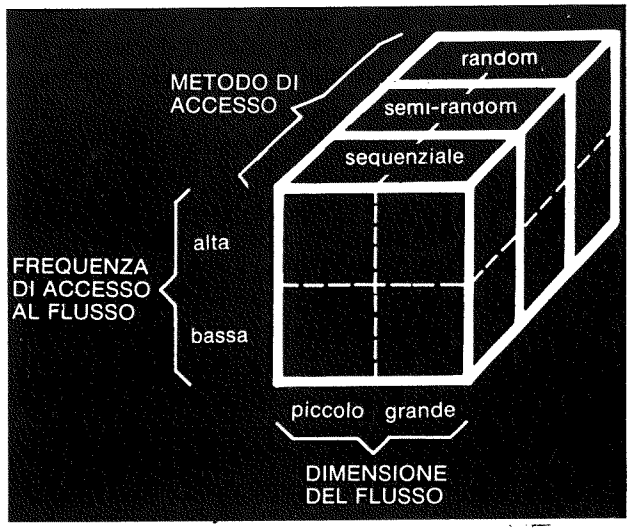


Fig. 5 - Stati del flusso.

L'AMMS deve decidere dove va memorizzata l'informazione in memoria ausiliare e come ottenere l'accesso ad essa. Questo combina i vantaggi dell'approccio distribuito per i flussi elaborati sequenzialmente, mentre predispone l'informazione per evitare la congestione creata quando il numero di stazioni di lettura/scrittura della MSF è limitato. Esso può anche limitare l'accesso al mezzo MSF predisponendo automaticamente i flussi usati frequentemente. Così le applicazioni più interessanti di entrambe le implementazioni, distribuita e gerarchica, possono essere incluse in un solo sistema.

I problemi nella definizione di strategie dell'AMMS per implementazione mista sono simili a quelli delle versioni più semplici. Cioè la strategia di distribuzione deve determinare dove devono essere memorizzati i flussi nella memoria ausiliare e la strategia gerarchica deve determinare l'algoritmo di caricamento, l'algoritmo di allocazione e l'algoritmo di riallocazione. La strategia distribuita è grandemente semplificata se tutte le informazioni del sistema sono memorizzate permanentemente nella MSF. Una caratteristica chiave è uno stretto legame tra la strategia di distribuzione e l'algoritmo di riallocazione; essi controllano entrambi i movimenti del flusso nella memoria di massa, l'uno dalla memoria principale, l'altro nella predisposizione. La strategia gerarchica deve anche stabilire come ottenere accesso all'informazione, se attraverso la gerarchia predisponendo il disco, o attraverso l'accesso diretto alla MSF. I fattori coinvolti in questa decisione sono le dimensioni del flusso, la frequenza di accesso al flusso, se i flussi sono sequenziali o non sequenziali, gli ingorghi sui meccanismi di accesso alla memoria di massa e gli ingorghi del disco.

5 - Un modello preliminare

L'approccio misto gerarchico - distribuito all'implementazione di un AMMS per MSF pone parecchi problemi interessanti. Ottenere accesso all'informazione direttamente nella memoria di massa dovrebbe semplificare alcuni degli aspetti della gestione gerarchica e permettere inoltre che una maggiore quantità di informazioni venga memorizzata nella memoria di massa. Per esempio, l'accesso diretto a un grande flusso sequenziale semplificherà la strategia di allocazione riducendo la congestione sul disco. Al contrario, predisponendo flussi non sequenziali più piccoli sul disco, la congestione sul meccanismo di accesso alla memoria di massa può essere ridotta.

I problemi che dovrebbero essere esaminati riguardano il luogo in cui un flusso dovrebbe essere memorizzato permanentemente e il modo in cui si dovrebbe ottenere l'accesso al flusso. Questo modello presuppone che la memorizzazione permanente dei dati del sistema sia distribuita su tutte le forme di memoria ausiliare. Esso presuppone anche che l'AMMS sia responsabile della determinazione del metodo di accesso. Proponiamo tre parametri di analisi: dimensioni, frequenza di accesso, e metodo di accesso, usando un semplice modello di flusso (Fig. 5).

Il metodo di accesso è random o semi-random o sequenziale. Il semi-random si applica a flussi altamente strutturati dove una tecnica random localizza un segmento del flusso che viene poi elaborato sequenzialmente. La frequenza di accesso del flusso dovrebbe in realtà includere una misura della localizzazione per mostrare quanto frequentemente si verifichino operazioni di predisposizione e quale è lo schema d'uso del flusso.

Una discreta suddivisione degli stati del flusso si ottiene dividendo le dimensioni del flusso e la frequenza di accesso al flusso come indicato dalle linee tratteggiate del diagramma di Fig. 4.

Si può individuare una strategia per ciascuna suddivisione per stabilire se il flusso in quella suddivisione deve risiedere sul disco residente, nella memoria di massa con accesso predisposto, o nella memoria di massa con accesso non predisposto. Per esempio, un flusso piccolo usato non frequentemente, ma che subisce un'intensa attività quando è usato, verrebbe tenuto nella memoria di massa e predisposto sul disco quando è usato. Per quelle suddivisioni per cui non c'è una

chiara strategia, la decisione di predisporre o no può essere basata sulle variabili del sistema come ad esempio la congestione del sistema di input/output della memoria.

6 - Sommario e conclusioni

L'introduzione di memorie MSF ha aperto all'industria del computer una nuova fase di sviluppo nel campo delle memorie di massa automatiche. Infatti questi dispositivi sono potenzialmente atti a realizzare memorie di capacità maggiore ad un costo inferiore e con accesso più rapido di qualsiasi altra forma precedente di memoria di massa automatica. Però, prima che questo potenziale possa essere effettivamente sviluppato, devono essere risolti parecchi problemi di architettura di sistema, tra cui problemi di configurazione dei dispositivi, tipi di dispositivi e configurazione del percorso dei dati tra i dispositivi.
Usare il disco e l'MSF in una combinazione mista distribuita - gerarchica è attualmente la configurazione più interessante. Strettamente legati all'architettura di memoria sono le strategie di gestione del sistema che determinano dove devono essere memorizzati i dati, come devono essere organizzati e quando devono essere spostati. Le soluzioni a questi problemi devono fornire una prestazione ottimale del sistema di memoria di massa, ma non a spese della sicurezza e della integrità dei dati memorizzati. Le MSF rappresentano il risultato iniziale di ricerche ingegneristiche e tecnologiche dirette ad ottenere grandi memorie di massa, a basso costo ed alta velocità. Tuttavia, se ciò permette la memorizzazione di grandi quantità di informazioni, non si otterrà alcun reale successo se le informazioni non potranno essere gestite in maniera efficace.

---

7 - Bibliografia

Si rimanda alla bibliografia contenuta in:
D.L. BOYD, "Implementing Mass Storage Facilities in Operating Systems", Computer, 11 n. 2, febbraio 1978, da cui è stato tratto il materiale del presente articolo.

Nuove tecnologie di visualizzazione per l'informatica

CLAUDIO DALMASSO
Ing. C. Olivetti & C.
Direzione Centrale della Ricerca
Ivrea

1 - Introduzione

L'uso di display nelle macchine per informatica è andato crescendo in questi anni, e con l'avvento delle tecnologie di programmazione interattiva il ricorso al display è entrato nell'esperienza di ciascun operatore.
Il tipo normalmente usato è il Tubo a Raggi Catodici (CRT), derivato dall'esperienza della televisione commerciale e adeguatamente modificato per rispondere alle esigenze di densità di informazione e chiarezza di presentazione che sono tipiche delle applicazioni informatiche.
L'ampia diffusione ne ha anche messo in evidenza alcuni aspetti negativi, e ha stimolato lo sviluppo di tecnologie alternative, nessuna delle quali tuttavia ha raggiunto oggi la completa capacità di sostituzione del CRT.
Tra gli aspetti negativi del CRT citiamo innanzi tutto, come intrinseci nella sua tecnologia costruttiva, il rilevante ingombro e la relativa pericolosità di esplosione.
La superficie curva della faccia di presentazione dei dati e le caratteristiche di emissione di luce dei fosfori costituiscono ancora elementi negativi, anche se meno gravi dei precedenti.
Le tecnologie alternative si sforzano anzitutto di rimediare a questi inconvenienti, ma finora non sono riuscite a giovarsi del principale vantaggio del CRT, consistente nella semplicità di controllo. Come vedremo in seguito, tutte le tecnologie alternative che si vanno proponendo richiedono sistemi di pilotaggio largamente più complessi del CRT, occorrendo a questo soltanto tre controlli per essere operato:
- scansione orizzontale del pennello elettronico
- scansione verticale del pennello elettronico
- modulazione (o innesco-soppressione) del pennello elettronico.

Una prima conseguenza di questo stato di cose la troviamo quando cerchiamo di confrontare sotto il profilo della convenienza economica due forme diverse di display e ci accorgiamo che non possiamo limitare il confronto all'elemento visualizzatore, ma dobbiamo spingere l'analisi attraverso la circuiteria elettronica che gli sta a monte fino a trovare un punto di interfaccia all'interno del sistema logico che rappresenti realmente il punto al quale le modifiche di progetto si possono arrestare.

2 - Generalità sul sistema display

Il "sistema display" rappresenta la connessione fra l'elaboratore elettronico e l'elaboratore umano. La sua principale funzione consiste nel trasformare il codice di informazione dall'una all'altra forma di elaborazione.
Gli sarà quindi chiesto sovente di operare in maniera bidirezionale, ovvero di trasformare il codice operativo di macchina in codice visivo adatto all'uomo, ma anche di essere in grado di ritrasformare una forma dal codice visivo in una sequenza logica che sia operativa per l'elaboratore.
I fattori che influenzano la composizione del "sistema display" dovranno quindi tenere conto di questo fatto oltre che ovviamente della tecnologia costruttiva del visualizzatore.
Un "sistema display" di regola sarà quindi composto nel seguente modo:
- una memoria elettronica, che lavora in codice binario e ha la funzione di trasformare la frequenza del flusso di informazioni logiche da quella più adatta all'elaborazione in macchina a quella necessaria per seguire il ritmo di lavoro tipico del visualizzatore.
- Alcuni circuiti di sincronizzazione che trasformano il flusso seriale di informazioni in una sequenza dotata dell'opportuno parallelismo richiesto dal visualizzatore (il grado di questo parallelismo varia secondo le tecnologie, dal caso più semplice di 1 per il CRT al caso massimo del numero di punti costituenti una riga, tipico dei più complessi display a matrice).
- Un generatore di caratteri che trasforma il codice logico nel codice di forma adatto alla lettura dello schermo.

- I circuiti elettrici che trasformano il segnale logico nei livelli di tensione richiesti dalla tecnologia del visualizzatore (su questo punto le nuove tecnologie sono in vantaggio rispetto al CRT in quanto richiedono tensioni di regola inferiori a 100 Volt contro le parecchie migliaia richieste da questo).
- Il visualizzatore vero e proprio.
- L'elemento interattivo o di "feedback" che sarà tipicamente una penna ottica o, nei casi più semplici, un cursore operato da tastiera.

3 - Generalità sulle nuove tecnologie

Nella ricerca di nuove tecnologie sostitutive del CRT vengono perseguiti numerosi obiettivi che per ora non riescono ad essere contemporaneamente soddisfatti da alcuna delle tecnologie note. Se in aggiunta agli obiettivi tecnici consideriamo l'aspetto economico, richiedendo un costo del nuovo "sistema display" non maggiore di quello del CRT, troviamo nelle tecnologie oggi esplorate delle grosse difficoltà di principio legate essenzialmente al maggiore ammontare di circuiteria elettronica che esse richiedono per il pilotaggio.
Gli obiettivi perseguiti nella ricerca di nuove tecnologie sono tipicamente:
- libertà di formato dello schermo. La struttura costruttiva del CRT favorisce i formati molto regolari, tipicamente quadrati;
- basso volume e perfetta planarità dello schermo;
- tensione di lavoro bassa, possibilmente compatibile con le normali tecniche dei circuiti integrati;
- strutture dello schermo predisposte per la presentazione di immagini di geometria molto semplice, ma con esigenza di grande precisione (tipicamente struttura a matrice).

Un aspetto particolarmente interessante e tipico di alcune nuove tecnologie è rappresentato dalla possibilità di presentare i dati in maniera passiva, ovvero per la semplice colorazione dello schermo senza emissione di luce. Alcuni aspetti del tipico problema di affaticamento dell'operatore davanti al CRT sarebbero risolti qualora si mettesse a punto una delle tecnologie che offrono la possibilità di presentare immagini passive. Non sempre per ora queste tecnologie consentono di controllare soddisfacentemente tutti i parametri ergonomici secondari, quali per esempio il contrasto e l'angolo visivo. Di regola i display realizzati con tecnologie passive richiedono poca energia per funzionare, consentendo quindi riduzioni di costo nella circuiteria di pilotaggio.
Due tecnologie principali stanno cercando la loro strada a livello industriale, attraverso una grande quantità di varianti. Una di queste è attiva, quella del plasma, e una è passiva, i cristalli liquidi.
Altre tecnologie tipicamente passive sono tuttora in fase di ricerca e ne accenneremo brevemente in seguito.

4 - I display a plasma

Il fenomeno fisico utilizzato nei display a plasma è quello della scarica elettrica nei gas. Come è noto que-

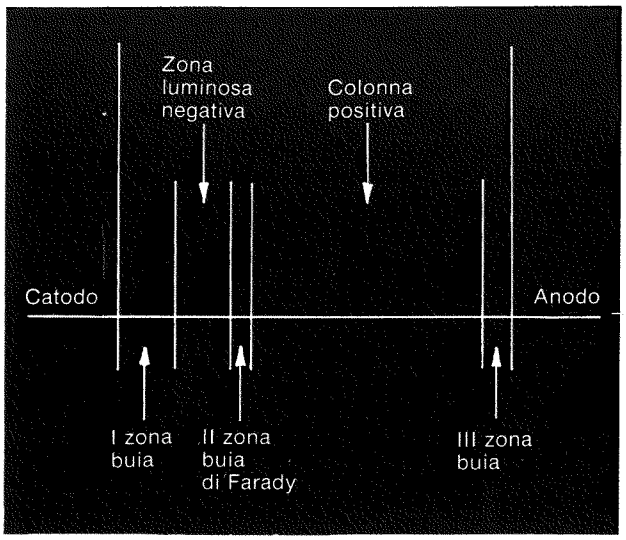
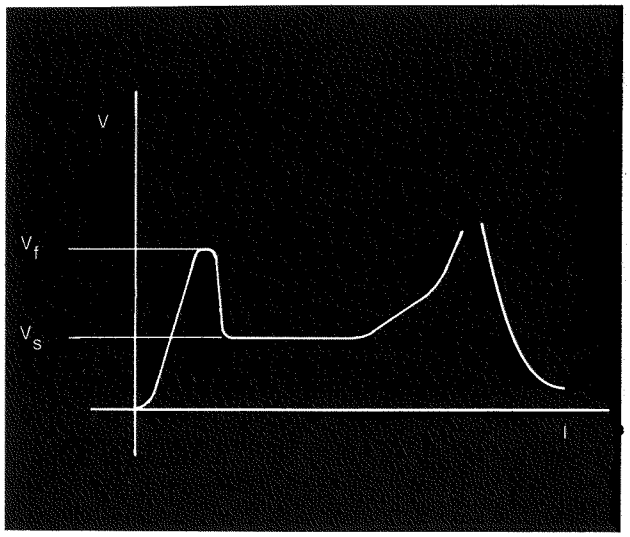


Fig. 1 - Struttura della scarica elettrica nei gas.

sta scarica presenta effetti luminosi divisi in due zone secondo lo schema in fig. 1. Di regola per realizzare i display a plasma il catodo e l'anodo vengono ravvicinati al punto che rimanga efficace la sola zona luminosa negativa.
I parametri critici della cella a plasma sono le tensioni di lavoro, i tempi di innesco e di decadimento della scarica e le frequenze luminose emesse. Senza entrare nella teoria della scarica nei gas diciamo semplicemente che questi parametri critici sono controllati nel progetto del display attraverso le caratteristiche geometriche della cella (tipicamente la distanza tra catodo e anodo), la scelta della miscela gassosa utilizzata (tipicamente neon-argon con tracce di mercurio) e la pressione del gas (tipicamente 200 Torr).
La caratteristica elettrica della scarica nei gas presenta tre regioni di funzionamento secondo quanto indicato in fig. 2 Si identificano due tensioni caratteristiche, V_f e V_s , che vengono rispettivamente utilizzate per accendere la cella e per mantenerla accesa.
Il meccanismo della scarica nei gas richiede, per un funzionamento efficiente, la presenza di una certa den-

Fig. 2 - Caratteristica corrente-tensione nella scarica elettrica nei gas.



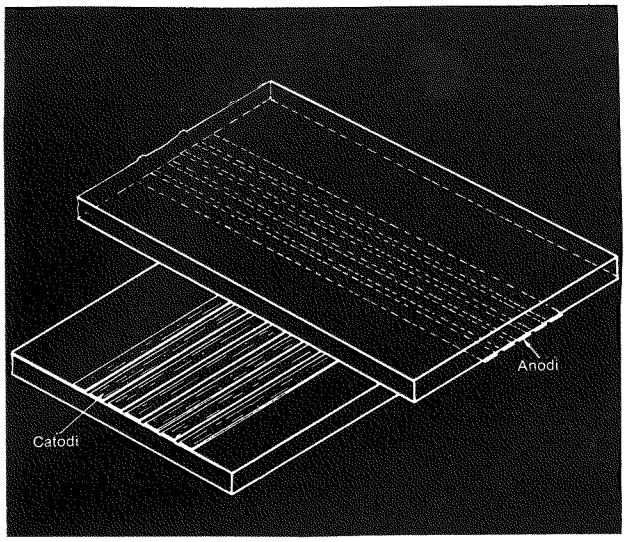
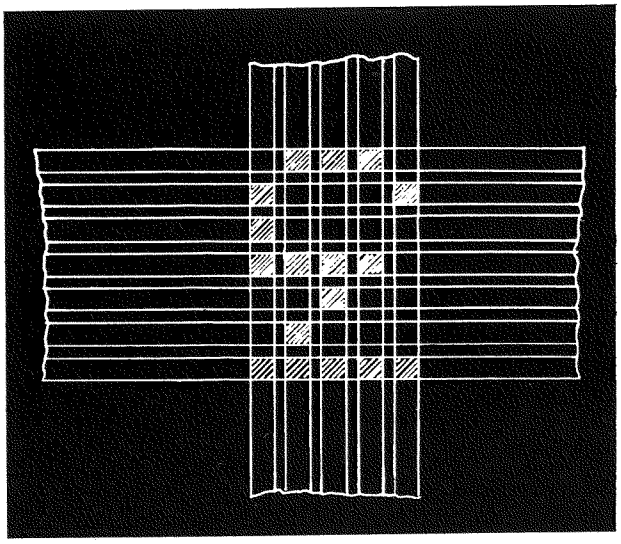


Fig. 3 - Schema generale del display a plasma.

sità di elettroni nel gas. Le difficoltà tecniche connesse all'introduzione di catodi caldi, capaci di emettere elettroni in quantità, in una struttura meccanicamente semplice quale deve essere quella di un display impongono di impostare il progetto dei display a plasma sulle tecniche dei tubi a catodo freddo. Di conseguenza occorre dedicare certe aree del display a celle con bassissima tensione di innesco nelle quali la scarica sia mantenuta continuamente in modo che esse servano da fornitori di elettroni e di fotoni per la parte visibile del display. Su questo principio, detto del "priming", si basano anche i tentativi di introdurre nel display a plasma un metodo di funzionamento per "autoscansione" attraverso il quale si tenta di ridurre il numero dei punti di accesso dell'alimentazione e quindi anche il numero dei circuiti di pilotaggio necessari.

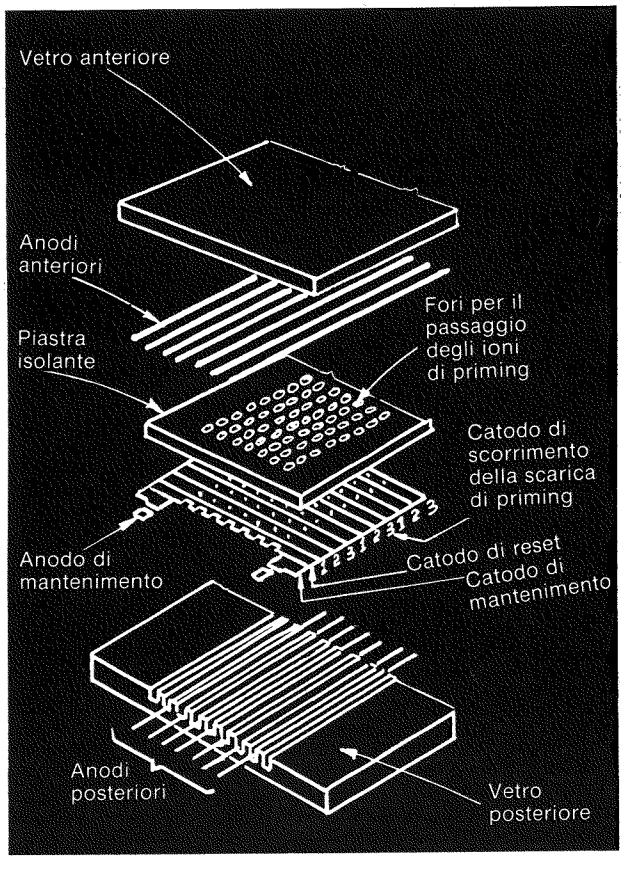
Di regola un display a plasma sarà costruito in base allo schema di figura 3 come una matrice di celle organizzate in modo che tutti i catodi possano essere disposti su una serie di linee su una faccia del display mentre tutti gli anodi possano essere disposti sulla faccia opposta su una serie di linee perpendicolari alle precedenti.

Fig. 4 - Rappresentazione matriciale dei caratteri. Esempio.



Le dimensioni tipiche delle celle sono di alcuni decimi di millimetro sia di spessore che di lato.
Il numero dei circuiti di pilotaggio richiesti è pari alla somma del numero delle righe e del numero delle colonne.
Per presentare, per esempio, un migliaio di caratteri in configurazione a matrice 5 x 7 (v. fig. 4) occorrono circa 500 circuiti di pilotaggio indipendenti.
Per tentare di ridurre il numero di circuiti necessari al pilotaggio del display si sono sviluppate delle tecniche che sfruttano il fenomeno del "priming" per generare all'interno del display uno stato di preeccitazione degli atomi e controllarlo in modo che la scansione in una direzione (per es. in direzione orizzontale) avvenga in modo automatico attraverso il controllo di pochi elettrodi (in questo caso colonne). Il numero di circuiti in questa situazione si riduce a poco più della metà del caso precedente.
La struttura del display ne viene sensibilmente complicata.
In fig. 5 diamo lo schema costruttivo di un display di questo tipo prodotto industrialmente. L'ordine di grandezza dei tempi caratteristici della scarica pone un limite alle dimensioni della matrice sia nel caso di accensione indipendente degli elettrodi sia nel caso di scansione. Questo limite si riflette in una lunghezza di riga difficilmente superiore a 40 caratteri. Un altro inconveniente presentato dalle tecniche di accensione per scansione è rappresentato da effetti di disturbo

Fig. 5 - Spaccato di un display a palsma a autoscansione.



visivo, che diventano particolarmente evidenti quando si sfruttano al limite le possibilità offerte dai tempi critici della scarica; ne consegue quindi che l'aspetto di questo tipo di display tende ad essere meno gradevole con l'aumento delle dimensioni. Per salire a dimensioni più grandi e avere una immagine più ferma si sono sviluppate delle tecniche di memoria capaci di stabilizzare la scarica nei punti in cui è stata innescata. La principale tecnica di memoria finora sviluppata si basa sull'effetto di condensatore che si può avere in una cella qualora gli elettrodi siano ricoperti di materiale isolante. In questo caso le cariche elettriche generate dalla scarica si accumulano sulle pareti isolanti della cella e, operando in corrente alternata, ne consentono la riaccensione a tensione particolarmente bassa. Diventa quindi possibile pilotare il display con due forme d'onda indipendenti, l'una destinata alle funzioni di accensione e spegnimento delle celle e l'altra destinata al mantenimento di uno stato esistente.

È possibile anche sfruttare questo meccanismo di mantenimento delle cariche su strati isolanti per provocare un fenomeno di autoscansione, rinunciando al fenomeno di memoria, raggiungendo lunghezze di riga sensibilmente superiori a quelle della tecnologia di scansione precedentemente citata. I display di questo tipo sono tuttora in fase di ricerca, mentre è possibile trovare a livello di realizzazione industriale display con memoria fino a 1000 righe per 1000 colonne di punti.

I display del tipo a elettrodi non protetti sono di uso corrente per alimentazione a coincidenza e in autoscansione su dimensioni fino a 250 righe per 250 colonne di punti. Essi presentano una relativa semplicità costruttiva, ma presentano le dimensioni limite di cui abbiamo parlato e sono soggetti a rischi di guasti per effetto di sputtering degli elettrodi. Questo tipo di display è infatti regolarmente operato in corrente continua ed è quindi difficile evitare che una parte del materiale degli elettrodi venga evaporato dalla scarica e vada a condensarsi sui vetri della cella o a provocare corto circuito tra gli elettrodi.

Per loro parte, i display con elettrodi isolati dal gas, che si operano in corrente alternata e quindi presentano rischi minori di sputtering, incontrano serie difficoltà costruttive nel controllo della uniformità del fattore di emissione elettronica secondaria dell'ossido che ricopre gli elettrodi.

Si stanno sviluppando nei laboratori industriali numerose soluzioni che adattano i display a plasma all'uso di una penna di luce per il controllo diretto dell'informazione da parte dell'operatore.

I display a plasma rappresentano oggi il tipo di display piatto industrialmente più diffuso e trovano il loro principale limite applicativo nell'alto costo della circuiteria di pilotaggio. Si può prevedere che il maggior sforzo di ricerca futura su questi display seguirà due linee di sviluppo: da una parte la diminuzione delle tensioni di pilotaggio attraverso lo sviluppo di materiali catodici a bassissimo potenziale di estrazione elettronica, dall'altra lo sviluppo di tecnologie integrate a semiconduttore capaci di sostenere le tensioni necessarie, e di tecniche di assemblaggio diretto dei semiconduttori sul pannello attivo.

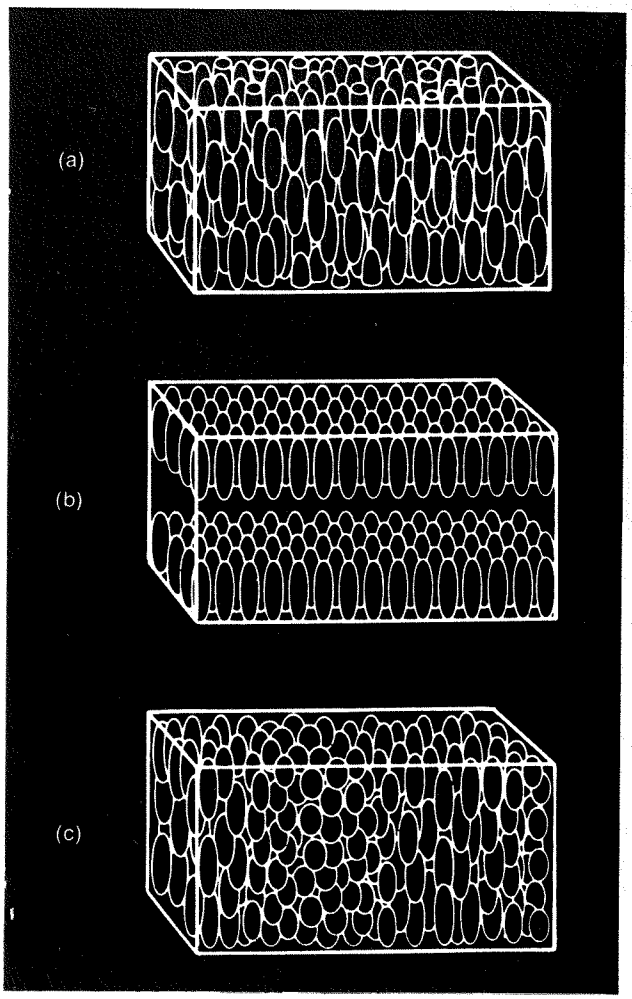


Fig. 6 - Tipologie principali di cristalli liquidi: a) Nematico, b) Smectico, c) Colesterico.

5 - I display a cristalli liquidi

Questo tipo di display è per ora molto diffuso nelle applicazioni di piccola superficie e per rappresentazioni numeriche a segmenti. Alcune notizie di realizzazioni in laboratorio inducono tuttavia a ritenere che se ne stiano sviluppando versioni adatte alla presentazione di dati in forma matriciale su dimensioni adatte alle applicazioni informatiche. A differenza dai display a plasma, che emettono luce, i display a cristalli liquidi hanno una presentazione tipicamente passiva, ovvero manifestano il loro effetto modificando una luce riflessa o trasmessa che ha origine all'infuori dell'elemento attivo del display.

Questa categoria di display si basa su una classe di materiali organici caratterizzata dalla forma molto allungata delle molecole e da particolari asimmetrie nel momento di dipolo elettrico delle molecole stesse. I fenomeni noti e sfruttati per la visualizzazione sono parecchie decine e si conoscono applicazioni a display che sfruttano come fonte della commutazione un campo elettrico, un campo magnetico, una radiazione termica.

In fig. 6 diamo uno schema dei tre ordinamenti molecolari più noti e più caratteristici dei cristalli liquidi. Questi ordinamenti molecolari rappresentano una fase

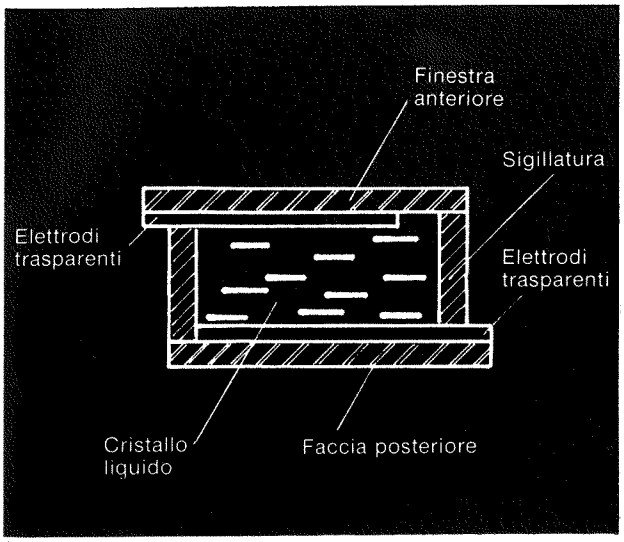


Fig. 7 - Cella di display a cristallo liquido a scattering dinamico.

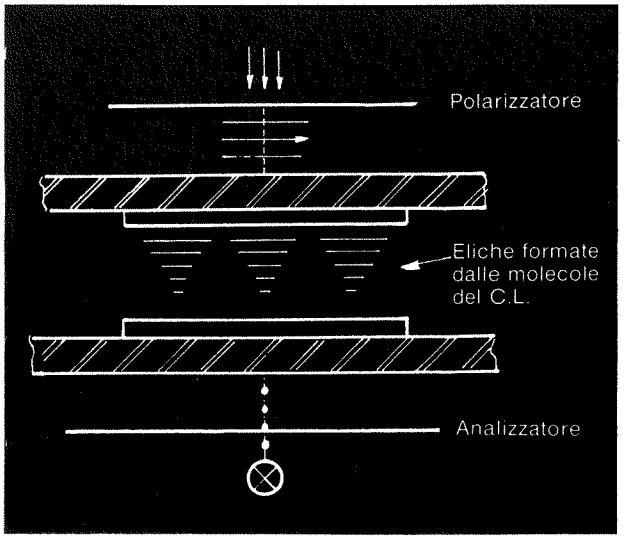


Fig. 8 - Cella di display a cristallo liquido nematico avvolto.

termodinamica della materia nel senso proprio della parola e sono stabili entro limitati intervalli di temperatura normalmente compresi tra una temperatura al di sopra della quale la sostanza è allo stato liquido (quale normalmente inteso) e un'altra temperatura al di sotto della quale la sostanza è allo stato solido. Lo stato di cristallo liquido mostra le caratteristiche macroscopiche di un liquido ma a livello molecolare possiede un grado di ordine quale normalmente si riscontra nei solidi. Alcuni effetti sfruttati per la realizzazione dei display si basano sulla esistenza di differenti fasi di cristallo liquido ordinate entro l'intervallo di temperatura citato. Nelle applicazioni finora diffuse l'effetto sfruttato è quello del riordinamento molecolare conseguente alla applicazione di un campo elettrico agli estremi di una cella che contenga un cristallo liquido (vedi fig. 7).

Le tre classi principali di cristalli liquidi danno luogo a effetti diversi. Il materiale nematico dà luogo alla maggior parte degli effetti finora utilizzati.

Uno di questi è il fenomeno di diffusione dinamica, secondo il quale la luce viene diffusa dalle zone attraversate da corrente a seguito della applicazione di un campo elettrico, nelle quali si induce un movimento turbolento delle molecole.

L'effetto di diffusione dinamica è caratterizzato sul piano applicativo da una curva di sensibilità alla tensione applicata, che non presenta un effetto di soglia ragionevolmente utilizzabile per la realizzazione di matrici, e da tempi di eccitazione e di decadimento dell'ordine di 100 millisecondi.

Le sue applicazioni più tipiche saranno quindi da ricercare nella presentazione di pochi caratteri a segmenti, il cui aspetto è quello ben noto di un carattere bianco brillante su fondo neutro.

Lo stesso materiale nematico può essere allocato nella cella con un tipo di ordinamento che ricorda l'ordinamento spontaneo delle molecole del materiale colesterico.

Questa struttura, raffigurata in fig. 8, è chiamata "nematico avvolto". Presenta fenomeni di soglia più netti che non nel caso precedente, tempi caratteristici leggermente minori, tensioni di lavoro dell'ordine del Volt (contro la decina di Volt del caso precedente), e possiede quindi una buona potenzialità per applicazioni di carattere matriciale. Essa è favorita anche dall'assenza di correnti all'interno del liquido.

La stabilità del materiale e la conseguente vita del display ne vengono largamente favorite.

La cella costruita secondo il principio del nematico avvolto viene osservata attraverso due polarizzatori, e quindi l'aspetto del carattere è nero su fondo grigio neutro. Il contrasto è molto migliore che nella diffusione dinamica, ma l'angolo visivo è minore e in alcune realizzazioni può essere abbastanza critico.

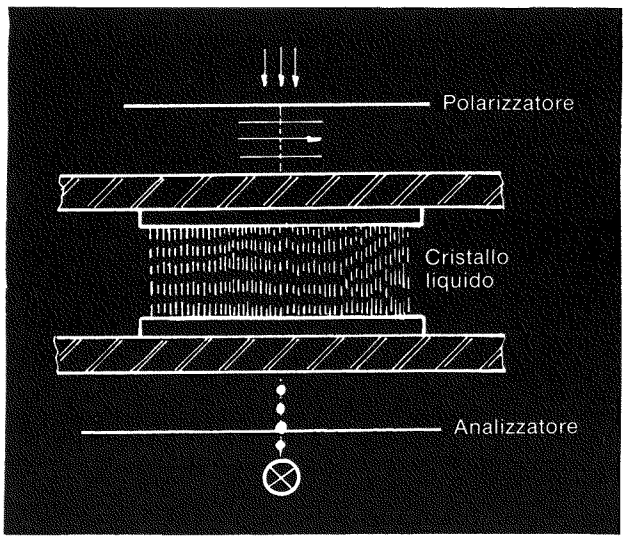
L'effetto ottico è dovuto al cambiamento di allineamento delle molecole del cristallo liquido in seguito all'applicazione del campo elettrico. Riferendoci alla fig. 8 osserviamo che all'atto della costruzione della cella le molecole vengono allineate parallelamente alle facce della cella (mediante un opportuno trattamento delle sue superfici interne), ma disposte in modo che nella direzione dello spessore della cella si formi un'elica che le ruota di 90°. In questa situazione l'attività ottica delle molecole è tale che il piano di polarizzazione della luce incidente viene ruotato di 90°, e ci è consentito quindi di rendere la cella completamente trasparente o completamente opaca semplicemente orientando la coppia polarizzatore-analizzatore in modo opportuno.

L'applicazione del campo elettrico agli estremi della cella allinea le molecole perpendicolarmente alle facce della cella, cancellando in tal modo l'attività elettrottica dell'area selezionata. In tale area l'effetto dei polarizzatori sarà quindi visto in modo complementare all'effetto delle aree circostanti.

Un fenomeno analogo si realizza disponendo (sempre mediante un appropriato trattamento delle facce interne della cella) un materiale nematico in modo che le molecole siano, allo stato di riposo, allineate perpendicolarmente alle facce della cella e assumano una posizione leggermente inclinata quando sollecitate dal campo elettrico. Lo schema di massima è mostrato in fig. 9. Il materiale adatto per presentare questo feno-

meno differisce da quello del fenomeno nematico avvolto in quanto alle molecole si richiedono caratteristiche dielettriche diverse.
La luce, sempre polarizzata, che attraversa la cella subisce in questo caso uno sfasamento proporzionale all'angolo formato dalle molecole rispetto alla posizione di riposo e allo spazio percorso. I tempi di commutazione legati a questo modo di lavorare possono essere resi particolarmente brevi, e la soglia di funzionamento particolarmente nitida, aggiustando lo spessore della cella in accordo con le caratteristiche elettriche delle molecole. È possibile anche graduare il ritardo di fase della luce in modo da generare colori interferenziali controllabili attraverso il campo elettrico applicato. Questo fenomeno appare, in linea di principio, molto promettente per la realizzazione di matrici di discrete dimensioni, ma presenta alcune difficoltà tecnologiche di un certo rilievo connesse alla realizzazione di spessori di cella molto piccoli (dell'ordine del micron) e uniformità controllate con estrema precisione.
Gli effetti che si presentano nei cristalli liquidi smectici e colesterici sono oggetto di ricerca, e per quanto abbiano dato risultati interessanti (tra cui per es. effetti di memoria) non risultano per ora aperti ad applicazioni industriali.
Abbiamo detto che i materiali che costituiscono i cristalli liquidi sono di natura organica. Nella loro forma più generale si tratta, come indicato in fig. 10, di due anelli benzenici legati tra di loro da una molecola scelta in una gamma opportuna e terminati con due gruppi scelti anch'essi entro opportune famiglie. Questi materiali sono stati oggetto di lunghe ricerche destinate ad aumentarne la stabilità, che oggi ha raggiunto un limite soddisfacente di almeno 30.000 ore quando la cella sia correttamente utilizzata.
A questo riguardo due sono i punti delicati che non in tutte le realizzazioni sono perfettamente risolti.
Il primo è legato al modo di funzionamento e alla perfezione costruttiva degli elettrodi, che, se permettono il passaggio di correnti elettriche entro il liquido, ne provocano la decomposizione elettrolitica, mentre il secondo è legato alle tecniche di sigillatura, dove si è a volte costretti a introdurre collanti o materie plastiche che possono interagire chimicamente col liquido qualora rimangano residui non perfettamente polimerizzati.

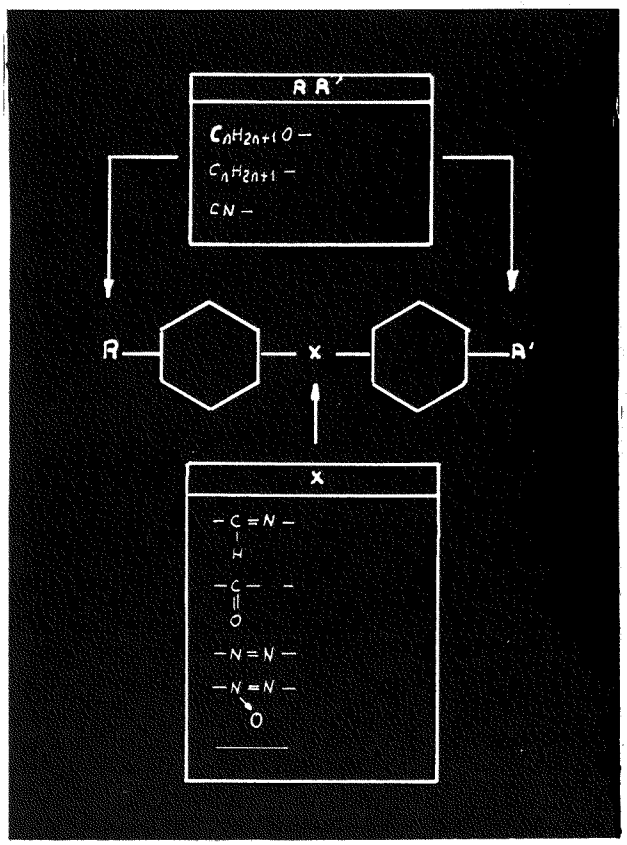
Fig. 9 - Cella di display a cristallo liquido a distorsione di fasi allineate.



6 - Le tecnologie in fase di ricerca

La ricerca sui display piatti è molto attiva e percorre strade tra di loro molto diverse, sia a livello tecnologico che a livello di ricerca di base sui materiali.
Sul piano tecnologico possiamo citare ricerche su tubi a raggi catodici di ridotto spessore ma di struttura alquanto complicata in quanto fanno ricorso a un reticolo di catodi distribuito su tutta la superficie del display.
Sono anche in sviluppo visualizzatori a particelle magnetiche costituiti essenzialmente da sospensioni di particelle sferiche magnetizzate la cui superficie è colorata per metà in nero e per metà in bianco.
Questo tipo di dispositivo sembra presentare interessanti potenzialità, ma deve ancora essere portato a un livello di complessità ragionevole prima di essere industrialmente interessante.
Un'altra tecnologia che ha avuto qualche sviluppo, e di cui si è recentemente parlato come possibile applicazione anche alle matrici, è quella dei visualizzatori elettroforetici: si tratta di una cella contenente un liquido opaco nel quale stanno in sospensione particelle colorate elettricamente cariche.
L'applicazione di un campo elettrico tra le facce della cella provoca il trasporto di queste particelle sulla superficie anteriore del display, e quindi la loro visibilità.

Fig. 10 - Struttura schematica delle molecole dei cristalli liquidi.



I tempi tipici di funzionamento con questa tecnologia sono lunghi, ma l'effetto intrinseco di memoria li può rendere ancora accettabili.
Altre due tecnologie da lungo tempo note, quella dei fosfori elettroluminescenti e dei fosfori catodoluminescenti stanno rivivendo, almeno per applicazioni particolari, una seconda giovinezza.
Nel campo delle ricerche a livello dei materiali si presentano come molto interessanti sotto l'aspetto del contrasto e della leggibilità i display elettrocromici, che presentano una immagine praticamente indistinguibile da una pagina stampata. L'effetto che si utilizza è la creazione di una banda di assorbimento ottico per effetto di un campo elettrico o di una corrente ionica.
Si conoscono effetti nei solidi, dove un elettrone intrappolato in una vacanza anionica può dare un centro di colore, o una banda di assorbimento può essere creata per trasferimento di carica tra centri di impurezza in un isolante.
Un'altra classe di questi fenomeni è studiata nei liquidi, o attraverso reazioni chimiche tipo indicatori di pH o attraverso ossidazioni-riduzioni controllate. Questi effetti presentano di regola una memoria propria. I principali problemi ancora da risolvere riguardano:
- vita e affidabilità dei materiali;
- tempo di commutazione, che per ora non scende sotto la decina di millisecondi;
- scarsa soglia, che rende difficile l'indirizzamento a matrice.

Un'ultima categoria di effetti di potenziale interesse si ritrova nelle ceramiche ferroelettriche. Sono aggregati di monocristalli (di dimensioni comprese entro 0,5 - 40 μm), caratterizzati da una polarizzazione netta spontanea che può essere invertita o riorientata per effetto di un campo elettrico applicato. Condizione essenziale perchè si presenti l'effetto ottico è che gli ioni nella cella unitaria siano disposti secondo una struttura che non presenta simmetria centrale.
Sono noti parecchi effetti, alcuni molto veloci, alcuni con memoria, alcuni con comportamento non lineare e conseguente possibilità di approccio matriciale.
Le principali limitazioni di questa classe di display derivano dalla tecnica di preparazione dei materiali (sinterizzazione) che consente per ora la preparazione di campioni di piccole dimensioni lineari e spessori non sufficientemente piccoli da portare le tensioni di pilotaggio entro limiti interessanti.

7 - Conclusione

La situazione delle nuove tecnologie di display è quindi sostanzialmente immatura, anche se sul piano ergonomico più di una tecnologia si presenta come promettente; considerazione dominante è che, se un grande sforzo di ricerca è rivolto a trovare nuovi materiali e nuovi fenomeni, tuttavia il massimo obiettivo pratico nella maggior parte dei casi è limitato all'ottenimento di una soglia sufficiente a consentire il funzionamento matriciale. Per molti settori applicativi questo limite è sufficiente, e possiamo facilmente prevedere che per i formati piccoli, tipicamente qualche decina di caratteri, o molto elongati, tipicamente da una a quattro righe contenenti fino a 80 caratteri ciascuna, le tecnologie del plasma, del fosforo catodoluminescente e dei cristalli liquidi vedranno allargarsi il loro mercato come sostitutivi del CRT, e anche in applicazioni per le quali un display non era stato finora preso in considerazione.
I problemi tipici dell'informatica riguardano però display di dimensioni nettamente superiori, da 480 a 5.000 caratteri.
In questi limiti il costo del "sistema display" con soluzione matriciale è ancora lontano dall'essere competitivo con il costo di un gruppo CRT, sia pur di tecnologia raffinata e altissima banda passante.
Il problema può essere attaccato in due direzioni:
- la tecnologia costruttiva del display può essere sviluppata in modo da consentire la massima integrazione della circuiteria di pilotaggio e la più drastica semplificazione strutturale;
- le proprietà dei materiali possono essere studiate in modo da generare fenomeni di propagazione molto veloce di stati preparatori, che adempiano all'interno del materiale allo stesso compito della scansione del pennello elettronico nel CRT.

Ricerche sono certamente in corso nella prima direzione, mentre sarebbe auspicabile che Enti di Ricerca debitamente interessati ai fenomeni fondamentali della materia volgessero il loro interesse al problema del secondo tipo.

8 - Bibliografia

Autori vari:
Tecnologie dei Sistemi di Informatica Olivetti, n. 9, "Tecnologie per visualizzatori alfanumerici". Luglio 1978.

Autori vari:
AEI - Tavola rotonda sui nuovi orientamenti nel campo dei trasduttori elettro-ottici per la ripresa e la presentazione delle immagini. Como - Sett. 1977.

Rivista professionale sui display: Information Display, the official journal of the Society for Information Display, mensile.

B. G. Couwenbergh et al. - Liquid Crystal Displays, Applicazioni Componenti Elettronici Philips, Vol. XV, 1978 n. 6, p. 50 e Vol. XV, 1978 n. 7-8, pagina 79.

La protezione fisica dei centri di elaborazione dei dati

ROBERTO MIGLIAVACCA
Honeywell S.p.A.
Direzione Automazione degli Edifici

1 - Introduzione

Una problematica generale del settore informatico è costituita dalla *sicurezza dei dati.* In questo ambito, un tema specifico è rappresentato dalla *sicurezza fisica dei centri di elaborazione dei dati.* Questo aspetto costituisce l'oggetto del presente articolo, mentre il tema più generale della sicurezza dei dati sarà materia di un articolo successivo.
L'importanza della protezione dei centri di elaborazione non ha bisogno di essere sottolineata. Nella società attuale, infatti, ogni tipo di attività - dalle industrie ai servizi di ogni tipo - è ormai strettamente dipendente dall'elaboratore. Proteggere queste macchine da danneggiamenti casuali o dolosi risulta quindi un obiettivo di importanza vitale.
La protezione fisica di un centro elaborazione dati si realizza con due tipi di mezzi:
- Passivi: quali caveau, porte corazzate, vetri antiproiettili, inferriate, sbarramenti in genere.
- Attivi: quali sensori elettronici di diverso principio di funzionamento e tra loro combinati.

Solo la combinazione degli elementi succitati, attivi e passivi, unitamente ad adeguate procedure operative, può garantire un sufficiente grado di sicurezza, così come la predisposizione di diversi livelli concentrici di protezione.
In questo articolo ci occuperemo della protezione fisica attiva.

2 - Il sistema di sicurezza

È opportuno prima di addentrarsi nella descrizione della architettura di un sistema di sicurezza identificare chiaramente quali obiettivi si vogliono raggiungere, ovvero da che cosa ci si vuol proteggere.
Si possono infatti temere o presumere diversi pericoli:
- L'incendio
- Il furto
- Il sabotaggio

I sistemi di protezione da porre in atto si differenziano tra loro a seconda del pericolo presunto; altro elemento importante è conoscere quale tipo di presidio fisso (guardiania) esiste all'interno dell'edificio per poter determinare i tempi e le forze di intervento in caso di allarme.
Nel caso specifico dei Centri di Elaborazione dei Dati si debbono presumere come potenziali pericoli sia l'incendio che il sabotaggio, mentre il furto, almeno per il momento, è un pericolo scarsamente temuto.
Abbiamo quindi appurato che il sistema di sicurezza ha lo scopo di proteggere il CED da incendio e sabotaggio; a questo punto però è necessario stabilire esattamente quali parti dell'edificio si devono proteggere.
L'attuale tendenza, soprattutto delle banche, è di decentrare il CED, in un edificio separato dalla sede centrale; questi edifici vengono comunemente denominati "Centri Servizi" e molto spesso sono situati in zone periferiche alla città; inoltre nello stesso edificio, unitamente alle attività strettamente legate al CED, coesistono altri servizi bancari, in genere quelli che non svolgono un'attività con contatto diretto con la clientela.
Si potrebbe a questo punto suddividere l'edificio in diverse aree e stabilire per ciascuna di esse il tipo di protezione ed il grado di sicurezza desiderabile ovvero il livello di rischio.
Una suddivisione potrebbe essere quella ipotizzata nella tabella 1.
Come si può constatare esistono tre zone per le quali il livello di rischio è elevato, sia per l'incendio che per il sabotaggio e dove la protezione contro questi due rischi è indispensabile.

3 - La protezione contro il pericolo d'incendio

Anche per questo argomento è indispensabile fare subito una distinzione tra:
- Sistemi di rilevazione automatica d'incendio
- Sistemi di estinzione dell'incendio

Può sembrare una distinzione ovvia ma purtroppo esiste ancora una certa confusione tra i due sistemi,

anche perchè sono molto spesso tra loro interagenti (il sistema di estinzione).
Dalla Figura 1 (stadi di sviluppo dell'incendio) si nota come al primo stadio, nel quale l'eventuale danno o pericolo per le persone è ancora molto basso ed il tempo sufficientemente elevato, è possibile intervenire, anche con mezzi manuali (estintori portatili, ecc), per estinguere il principio d'incendio limitando quindi i danni.
Il fattore tempo è sempre l'elemento determinante; si pensi ad un incendio durante le ore di non occupazione di un edificio e senza presidio fisso, ove si debba intervenire dall'esterno; ne consegue che il sistema di rilevazione incendio deve essere in grado di segnalare l'incendio al primo stadio.
Riferendoci ad una applicazione di tipo civile, quale è un Centro Servizi, possiamo dire che il sensore adatto a rilevare l'incendio nel suo primo stadio di sviluppo, è il tipo a doppia camera di ionizzazione.
Il secondo stadio di sviluppo, presenza di particelle di fumo, è normalmente coperto dai sensori ottici o a fotocellula. Per il terzo stadio di sviluppo si utilizzano i sensori termici o termovelocimetri ed infine per il quarto stadio sono impiegati i sensori ad infrarossi o ultravioletti.
Come già detto è importante rilevare un incendio al primo stadio, ma ciò non è sempre possibile in quanto molto dipende dal tipo di materiale presente; i materiali plastici o sintetici in genere, la moquette ad esempio, sono interessati ai primi due o tre stadi, mentre la benzina o simili in pratica non hanno il primo e secondo stadio, ma immediatamente si trovano al terzo o meglio al quarto stadio.
Per concludere quindi la scelta del sensore più adatto deve tendere a rilevare un incendio il più presto possibile in funzione dei materiali presenti in un certo ambiente.
Riferendoci ancora alla tabella 1 si ritiene necessario proteggere gli uffici, anche se il danno causato da un incendio potrebbe essere non molto elevato, ma potrebbe invece essere causa indiretta di elevati danni alle aree ove il rischio è classificato elevato, in caso di propagazione dell'incendio stesso.
Fin qui si è visto il sistema di rilevazione automatica che ha il compito di segnalare l'incendio; a questo punto entrano in gioco i sistemi di estinzione che a loro volta si suddividono in:

- Manuali

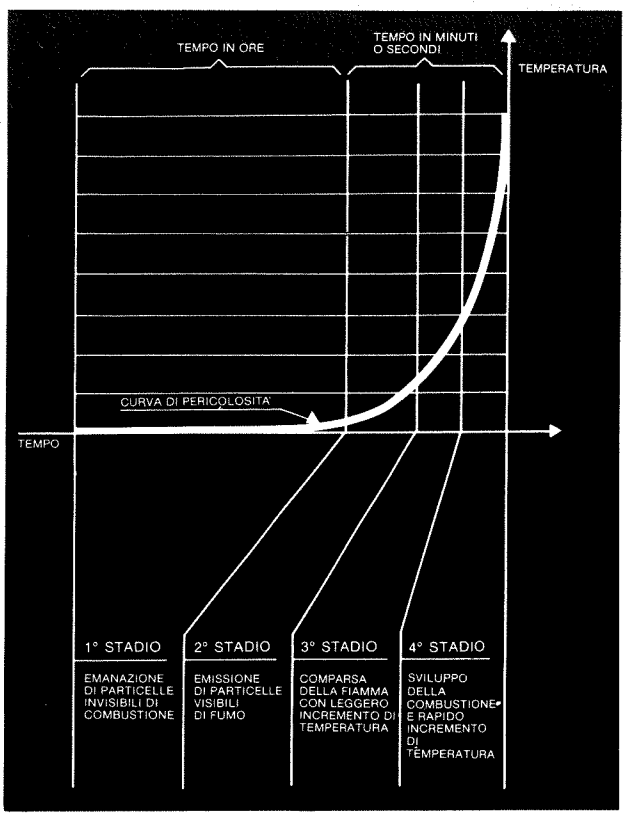


Fig. 1 - I quattro stadi di sviluppo dell'incendio.

- Automatici

Tra i manuali rientrano gli estintori portatili a schiuma e le manichette antincendio ad acqua; tra gli automatici: gli sprinkler, il CO_2 ed il gas Halon.
I campi di applicazioni dei diversi sistemi di estinzione variano a seconda delle aree e dei materiali, ferma restando la prescrizione dei Vigili del Fuoco; di norma nelle zone uffici si impiegano sistemi di estinzione manuali, nelle zone archivi possono esserci sprinkler (con cautela in quanto molto spesso il danno causato dall'acqua è superiore a quello dell'incendio) o gas Halon, per i locali degli impianti tecnologici dipende dal loro contenuto, gli impianti elettrici, come cabine trasformazione ecc, utilizzano il CO , mentre negli impianti di condizionamento, centrale termica e frigorifera, molto spesso si utilizzano sistemi manuali con estintori a schiuma.
Infine per la sala macchine e la nastroteca l'unico estintore usato è il gas Halon.
Lo scarico della carica estinguente di gas Halon è nor-

Tabella 1 - Livello di rischio/tipo di protezione

AREA	INCENDIO		SABOTAGGIO	
	Livello rischio	Protezione	Livello rischio	Protezione
Sala macchine	elevato	indispensabile	elevato	indispensabile
Nastroteca o Caveau Nastri	elevato	indispensabile	elevato	indispensabile
Zona uffici	normale	necessaria	basso	facoltativa (*)
Impianti Tecnologici	normale/elevato	indispensabile	elevato	indispensabile
Archivi	normale	indispensabile	basso	facoltativa (*)

(*) Eventuale sistema antintrusione per uffici o archivi ove vi siano documenti riservati.

malmente comandato in modo automatico dal sistema di rilevazione d'incendio; considerato il costo elevato di questo gas, è necessario prendere alcuni accorgimenti, affinchè un falso allarme non provochi uno scarico inutile del gas stesso.
La soluzione comunemente adottata è quella di realizzare una doppia maglia di rivelatori, sia a soffitto che sotto pavimento, ove esista, in modo tale che il comando dello scarico del gas sia effettuato solo quando due rivelatori della stessa zona sono contemporaneamente in allarme, e con una opportuna temporizzazione consenta agli occupanti di abbandonare la zona prima che lo scarico avvenga.
L'argomento richiederebbe una trattazione più estesa o particolareggiata e quindi queste note hanno il solo scopo di fornire alcune indicazioni orientative.

4 - La protezione contro il sabotaggio

Prima di esaminare quali misure porre in atto e quale sistema di protezione adottare è necessario identificare da chi e come può essere tentato il sabotaggio:
- da elementi esterni, sia durante il giorno che di notte, nei modi purtroppo conosciuti quali il sequestro di guardie o di personale del Centro;
- da elementi esterni, ma visitatori abituali quali fornitori, addetti alle pulizie o ai servizi di mensa, da impiegati che però svolgono la loro attività in altre sedi;
- da elementi interni che svolgono la loro attività nel Centro stesso.

Un altro importante fattore che determina il grado di sicurezza ottenibile è la struttura o la dislocazione dello edificio ove il Centro è stato installato; da questo punto di vista la miglior soluzione è rappresenta da un edificio appositamente progettato allo scopo e dove è possibile realizzare le opportune protezioni fisiche passive. Molto spesso invece vengono utilizzati edifici esistenti che difficilmente si adattano alla realizzazione di un Sistema di protezione affidabile.
Ciò premesso, ne consegue che la soluzione del problema, in conseguenza delle variabili in gioco, può variare notevolmente da edificio ad edificio e pertanto si possono dare solo delle indicazioni di carattere generale.
- *Un solo ingresso principale* sia per le persone che per i veicoli, con controllo visivo diretto oppure mediante telecamera e smistamento del traffico veicoli e pedonale. Ogni altro eventuale ingresso deve rimanere chiuso ed in caso di necessità aperto alla presenza di guardie e previo consenso dal Centro di Controllo (che sarà descritto più avanti), dopo opportuno riconoscimento della identità delle persone mediante telecamera.

Tutte le persone (dipendenti e visitatori) dovranno essere avviate all'ingresso dell'edificio.
- *L'ingresso dell'edificio* sarà munito di una doppia porta a trappola (quando è aperta la porta esterna deve essere chiusa quella interna e viceversa) telecomandata dal box di controllo, con struttura protetta, entro il quale risiederanno due persone per il ricevimento e la registrazione dei visitatori ed il riconoscimento degli impiegati.
- *L'organizzazione del traffico interno* è elemento indispensabile per poter garantire un livello minimo di protezione; a questo proposito è buona norma suddividere l'edificio in diverse aree con differenti livelli di accesso limitando l'ingresso alle aree stesse solo alle persone che operativamente hanno necessità di accedere ad esse.

Una possibile soluzione potrebbe essere:
- *zona verde* (uffici) per tutto il personale ed i visitatori.
- *zona rossa* (uffici CED) per il solo personale addetto al CED e a coloro che per il tipo di lavoro svolto hanno necessità di entrare in questa zona.
- *zona nera* per il solo personale addetto agli elaboratori ed alla nastroteca.
- *Il controllo di accesso,* mediante badge, quale mezzo per dirigere il traffico interno organizzato su diversi livelli crescenti di gerarchia per entrare nelle diverse zone.

Il primo controllo, corrispondente al primo livello, sarà all'ingresso dell'edificio; attraverso questo controllo dovranno transitare tutte le persone, impiegati e visitatori, muniti di badge.
Gli impiegati avranno un badge personale mentre i visitatori lo riceveranno, previa autorizzazione e registrazione, di volta in volta al box di controllo.
Ogni badge sarà codificato con il livello di accesso, i giorni della settimana e gli orari per ciascun giorno, per i quali è consentito accedere alle diverse zone.
La figura 2 rappresenta una proposta per l'ingresso dell'edificio.
- *Il controllo delle persone e degli oggetti,* laddove esista un alto fattore di rischio e si desideri avere un elevato grado di sicurezza, mediante "metal detector" e raggi X, rispettivamente per le persone e gli oggetti.
- *I sensori per la protezione attiva;* come già detto all'inizio, si deve partire dall'ipotesi di realizzare una serie concentrica di livelli di sicurezza, non è infatti pensabile di potersi affidare ad un solo sistema, se pur sofisticato.

Il *primo livello* di sensori sarà posto in corrispondenza della recinzione e consentirà di segnalare tempestivamente un'intrusione e di identificare il punto di ingresso. Per inciso è bene ricordare che un sistema di sicurezza ha il duplice scopo di rivelare un effrazione e al tempo stesso di far guadagnare più tempo possibile all'intervento delle guardie; da qui nasce l'esigenza di segnalare l'intrusione dal punto più esterno alla proprietà.
Un *secondo livello* di protezione sarà costituito dai sensori alle finestre ed alle porte d'ingresso all'edificio.
Un *terzo livello* sarà infine assicurato dai sensori a protezione di zone specifiche all'interno dell'edificio, quali caveau nastri, archivi ed impianti tecnologici.
A questo proposito è importante sottolineare come gli impianti elettrici al servizio degli elaboratori, debbano essere protetti, con controllo di accesso e sensori, in quanto un eventuale sabotaggio ad essi metterebbe il CED fuori servizio.
- *Il controllo di ronda* è un necessario completamento dei sistemi di sicurezza già descritti, in quanto realizza il controllo umano unitamente a quello automatico.

È altresì importante verificare che tutti i sistemi fin qui descritti, blocco degli accessi, porte, ecc. non costituiscano una trappola agli occupanti dell'edificio in caso

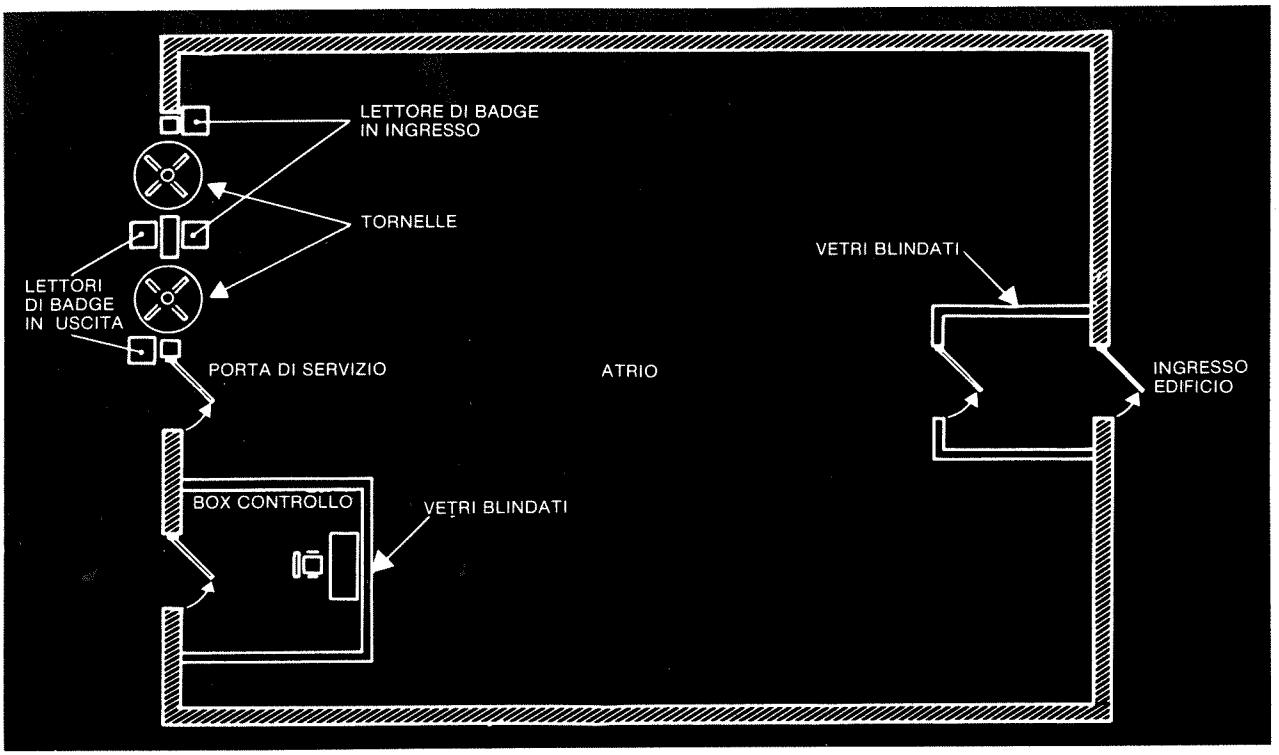


Fig. 2 Proposta per ingresso edificio.

d'incendio e quindi dovranno essere predisposte le uscite d'emergenza, pur se controllate con segnalazione d'allarme quando risultano aperte, tenendo presente che potenzialmente rappresentano dei punti deboli del sistema di protezione.

5 - Il centro di controllo

Tutti i sistemi di protezione fin qui descritti, rilevazione d'incendio, controllo di accesso, antintrusione e ronda costituiscono impianti autonomi, ma che tra di loro devono interagire; da qui la necessità di poter disporre di un unico Centro di Controllo che sia in grado di centralizzare e gestire i sistemi suddetti.

Ormai da anni la soluzione esiste ed è costituita dai sistemi computerizzati che uniscono alla semplicità d'impiego una grande flessibilità che consente di poter utilizzare questi sistemi per centralizzare e supervisionare non solo gli impianti di sicurezza, ma anche tutti gli impianti tecnologici compresi nell'edificio, quali quelli di condizionamento, riscaldamento, refrigerazione, gli ascensori, i gruppi elettrogeni, i convertitori, ciò allo scopo di garantire la piena efficenza degli impianti stessi, di assicurare il confort necessario negli ambienti e di poter programmare gli interventi di manutenzione.

Oltre a ciò questi sistemi, per mezzo di opportuni programmi standardizzati, sono in grado di realizzare considerevoli economie sui costi di gestione, anche del 30%, rendendoli pertanto un investimento profittevole dal momento che ripagano, con quanto è possibile risparmiare, il loro costo d'installazione nell'arco da 1 a 3 anni.

Infine i sistemi di centralizzazione provvedono anche alla registrazione scritta di tutti gli eventi che si verificano negli impianti da essi controllati, fornendo quindi una documentazione anche a posteriori particolarmente necessaria per gli impianti di sicurezza.

Questi stessi sistemi, mediante l'utilizzo delle linee trasmissione dati, possono centralizzare e controllare gli impianti di sicurezza e tecnologici di altri edifici. Nel caso specifico degli Istituti Bancari possono, ad esempio, centralizzare gli impianti di sicurezza delle diverse agenzie utilizzando in contemporanea le stesse linee di trasmissione dei dati bancari.

6 - Affidabilità

Riteniamo abbastanza evidenti i motivi espressi nel paragrafo precedente a proposito del vantaggio operativo di poter disporre di un unico strumento, il Centro di Controllo, che da solo sia in grado di gestire e controllare in modo integrato e semplice i diversi sistemi di protezione fin qui descritti, unitamente a tutti gli impianti tecnologici.

È però naturale porsi a questo punto la domanda riguardante il livello di affidabilità di un sistema di controllo quale quello descritto, anche dal punto della possibilità di sabotaggio delle informazioni.

Innanzitutto questi sistemi sono dotati di circuiti di autodiagnosi tali da rilevare eventuali guasti dei diversi componenti dell'intero impianto inteso come Unità Centrale di Elaborazione, linee di trasmissione, unità periferiche e sensori; inoltre, a garanzia degli eventuali sabotaggi, i sistemi utilizzano doppie linee di trasmissione dati con percorsi separati, doppia trasmissione delle informazioni con verifica bit per bit ed è anche possibile realizzare una codifica crittografica delle informazioni stesse.

7 - Conclusione

Allo stato attuale dell'arte questi sistemi di controllo integrato rappresentano quanto di più sofisticato ed affidabile, unitamente ad una grande semplicità d'uso, sia realizzabile allo scopo di ottenere un elevato grado di protezione fisica di un Centro di Elaborazione dei Dati.

Composizione di musica classica mediante elaboratore elettronico

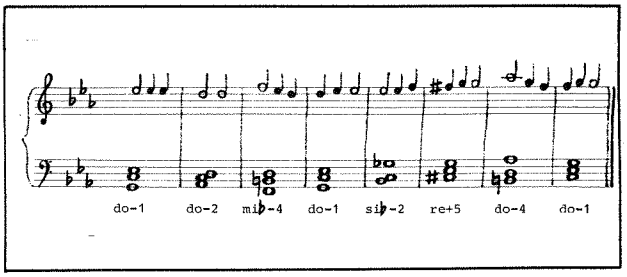
E. GAGLIARDO - P. FORNASARI
Istituto di Matematica
Università di Pavia

1 - Introduzione

Al sorgere di un nuovo ramo della matematica è assai difficile prevederne il ruolo ed i futuri sviluppi. Cercando tuttavia di rispondere alle domande che un recente esperimento di composizione di musica di stile classico mediante elaboratore elettronico ha suscitato, sembra ragionevole seguire le seguenti norme:
(A) Inquadrare questa ricerca nell'ambito di studi musico-cibernetici precedenti, se possibile.
(B) Diffondere la conoscenza di questo esperimento e dei suoi risultati, per permettere una discussione tra persone di diversa cultura.
Gli studi sull'acustica, sull'armonia e sulla melodia, che risalgono all'antichità e hanno sempre avuto un carattere almeno in parte matematico, hanno generato un complesso di norme di composizione facenti appello alla logica umana del compositore. Volendo render tali norme eseguibili da una macchina elettronica è innanzitutto necessario riformularle con un linguaggio esclusivamente matematico e, almeno allo stato iniziale di queste ricerche, assai semplificato. D'altra parte una volta stabilito questo linguaggio e codificati in esso numerosi brani di musica classica, un elaboratore elettronico programmato con i metodi della cosiddetta "intelligenza artificiale" può facilmente sviluppare complesse statistiche e scoprire strutture normative inconsciamente seguite dai grandi compositori del passato. La grande complessità di queste regole rende impossibile all'uomo la loro applicazione; è però possibile far eseguire i calcoli dall'elaboratore elettronico per poter esaminare quali composizioni musicali ne risultano.
È noto che le macchine elettroniche possono non solo "eseguire" brani musicali (simulando le vibrazioni dei vari strumenti) ma anche "comporre" musiche "ultramoderne", e la letteratura al riguardo è troppo vasta per essere citata.
È anche noto che il matematico R. Zaripov dopo molti anni di studi sull'argomento ha ottenuto un algoritmo mediante il quale un elaboratore elettronico può "comporre" melodie di stile non moderno (v. bibliografia).
L'esperimento riferito e brevemente descritto in questo articolo è stato programmato per esaminare la possibilità di comporre musiche di stile classico utilizzando elaboratori elettronici.

2 - Codificazione di brani musicali

Un linguaggio matematico per codificare un complesso fenomeno (non completamente conosciuto) in modo da rendere non illusorie ma significative le successive statistiche deve avere le due seguenti caratteristiche:
(a) Semplicità delle strutture con le quali gli eventi considerati essenziali vengono codificati.
(b) Raggruppabilità in classi di equivalenza (a livelli diversi) per gli eventi che sembrano avere caratteristiche analoghe.
Tenendo conto di questi requisiti, numerosi brani di musica classica sono stati codificati scegliendo come "unità" una successione di 8 battute, con ritmo4/4, con un accordo di 3 note per ogni battuta dell'accompagnamento, e priva di crome e di note più brevi della melodia.
Ad esempio uno dei temi della 7' sinfonia di Beethoven può così essere riassunto:



In questo brano l'accompagnamento è stato calcolato in base alle regole di armonia-melodia descritte nel seguito.

3 - Come l'armonia può essere insegnata a un calcolatore elettronico

Intendiamo identificare col nome di "nota" tutte le note le cui frequenze variano per un fattore 2^k con k intero relativo, cioè tutte le note di ugual "nome" appartenenti a qualsiasi ottava. Di conseguenza "identifichiamo" un accordo con i suoi "rivolti" (ad es. "do-mi-sol" con "sol-do-mi") e cioè, secondo l'osservazione (b) fatta precedentemente, rimandiamo ad un secondo tempo la scelta tra essi.

Le principali relazioni tra due "note" M, N, le cui frequenze verranno pure indicate con M, N, sono le seguenti.

Diremo che M, N sono tra loro "armoniose" se $N = 3 \cdot 2^k \cdot M$ (approssimativamente sulla "scala temperata") e cioè, seguendo le notazioni tradizionali, se le due note possono essere poste, con spostamenti di "ottave", a distanza di "quarta" o di "quinta" come ad esempio "do" - "sol". Si noti che questa relazione come anche quelle che defineremo in seguito è "orientata" ("sol" è l'armonica terza di "do" e non viceversa) ma allo stadio attuale della teoria non terremo conto di ciò, sebbene anche le statistiche suggerirebbero di iniziare a tenerne conto.

Diremo che M, N sono tra loro "melodiose" se $N = 15 \cdot 2^k \cdot M$ (approssimativamente) e cioè se le due note possono essere poste a distanza di "semitono" come "do" - "si".

Altre relazioni, meno rilevanti, sono quella "tonale": $N = 9 \cdot 2^k \cdot M$ (approssimativamente) e al tempo stesso : $M = 7 \cdot 2^h \cdot N$ (con peggiore approssimazione) e cioè se M, N possono essere poste a distanza di "tono" come "do-re", e quella "bitonale": $N = 5 \cdot 2 \cdot M$ (appr.) cioè se M, N possono essere poste a distanza di "terza maggiore" come "do-mi".

Disegnamo un reticolo di triangoli e rappresentiamo le note con i vertici facendo corrispondere a due consecutive tra le sei direzioni orientate che escono da ogni vertice le relazioni: "armoniosa" e "bitonale". I vertici di ogni traingolo con lato orizzontale inferiore (ruotando opportunamente la figura) corrispondono alle note di un "accordo maggiore". (Fig. 1-3)

Estendendo la reticolazione, come si vede facilmente, si viene a ricoprire una superficie "torica" sulla quale i problemi di armonia e melodia si studiano con maggiore facilità. (Fig. 4)

Diremo "tetraaccordo" un insieme di 4 "note" di nome diverso e tali che tra esse:

(I) Non figurino due note tra loro "melodiose" (cioè a distanza di semitono, a meno di ottave, come "do-si")

(II) Non figurino tre note consecutivamente "tonali" (come "do-re-mi").

Esistono 72 tetraaccordi e cioè 6 tipi diversi che possono essere "traslati", dioè spostati rigidamente lungo la tastiera del pianoforte, in 12 posizioni.

Questi 6 tipi, scegliendo i loro rappresentanti in modo da rimanere nell'area di do maggiore, possono essere così individuati: do-mi-sol-la (con notazione tradizionale "la" minore-settima), sol-si-re-fa (trad.: "sol" maggiore-settima), do-re-fa-lab (trad. "fa" dorico), si-re-fa-lab (trad. diminuito), sol-si-reb-fa (trad.: francese), sol-do-re-fa (di minor valore a causa della doppia relazione tonale: do-re, fa-sol).

Diremo "triaccordo" un insieme di 3 "note" che possa essere completato in un "tetraaccordo" (ammissibile) aggiungendo una opportuna quarta nota, in uno o più modi.

Questa definizione esclude gli accordi a 3 note che non soddisfano alle proprietà (I) e (II) sopraelencate, ed esclude anche gli accordi del tipo do-mi-sol ≠ di cui si trova un esempio nella 32' battuta della 7' sinfonia di Beethoven.

Tra i "triaccordi" hanno particolare importanza per le mutue relazioni "armoniosa" e "bitonale" delle loro note i 12 triaccordi "maggiori" (del tipo "do-mi-sol") e i 12 "minori" (del tipo "do-mib-sol"); grazie a queste 12+12 "tonalità" tutti gli altri triaccordi possono essere classificati nel modo seguente.

Dato un triaccordo si considerano i triaccordi maggiori o minori suoi "risolventi" cioè a cui esso risulta maggiormente collegato in base alle relazioni definite precedentemente. Ad esempio "sol-si-fa" è collegato al triaccordo maggiore risolvente "do-mi-sol" (infatti "do" è armonioso con "fa" e melodioso con "si, "mi" è armonioso con "si" e melodioso con "fa", "sol" è in comune e bitonale con "si"), ma è anche collegato al triaccordo maggiore risolvente "sol b - sib - reb" (infatti solb e sib hanno relazioni armoniose e melodiose con sol, si, fa). Questa seconda interpretazione con

Fig. 1 - Modulazioni "diatoniche" (spostamenti orizzontali) e "cromatiche" (spostamenti verticali) nella "Danza delle ore" di Ponchielli.

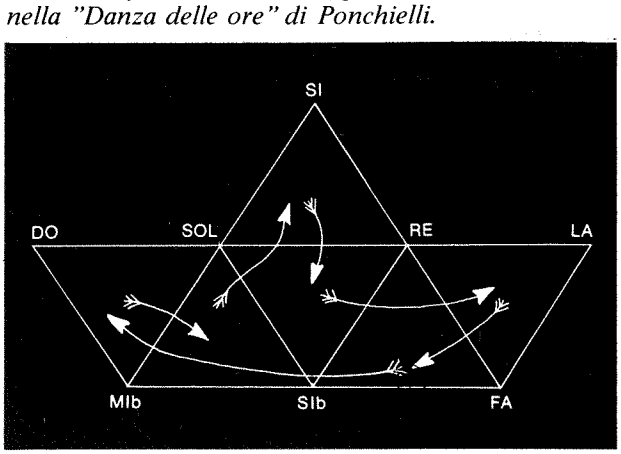
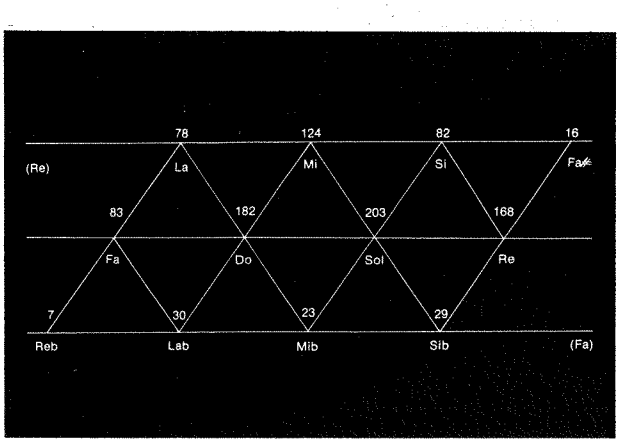


Fig. 2 - Frequenza di note (rappresentate come descritto in questo articolo) in brani di Schubert.



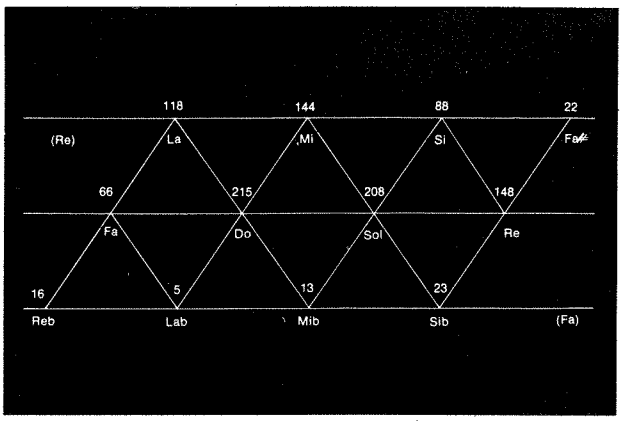


Fig. 3 - Una analoga statistica su brani di Schumann.

una traslazione di "tonalità" individua il triaccordo "reb-fa-si" che risolve in "do-mi-sol".

Diremo d'ora in poi "accordo" un "triaccordo" associato ad uno dei suoi "risolventi".

Conveniamo che un triaccordo "maggiore" o "minore" ammetta se stesso come unico "risolvente" e quindi abbia una sola interpretazione. "Gli "accordi" principali risultano 144, cioè si hanno 6 tipi per ciascuno dei 12 "risolventi" maggiori, e 6 tipi per ciascuno dei 12 "risolventi" minori. Questi 6+6 tipi si possono individuare elencando i 6 triaccordi che risolvono in "do" maggiore (che sono: "do-mi-sol" stesso, "sol-si-fa", "sol-re-fa", "si-fa-lab", "reb-fa-si", "do-fa-sol"), e i 6 triaccordi che risolvono in "do" minore (che sono: "do-mib-sol" stesso, "re-lab-do", "re-fa-do-", "si-re-lab", "lab-re-solb", "sol-do-re"). Osserviamo che sulla superficie torica di cui si è parlato i 6 tipi di accordi con risolvente maggiore hanno configurazioni simmetriche a quelle dei 6 tipi di accordi con risolvente minore.

La classificazione degli "accordi" ora introdotta consente di indicare con (I, J), $1 \leqslant I \leqslant 24$, $1 \leqslant J \leqslant 6$, l'accordo di tipo J che risolve nella "tonalità" I.(*)

Introduciamo ora il concetto di "scala di un accordo" nel modo seguente: Nel caso di un triaccordo maggiore o minore la scala è costituita dalle 3 note del triaccordo e dalle 4 note con esse "armoniose". Nel caso generale si considerano anzitutto:

(i) le 3 note del triaccordo,

(ii) le note che risultano armoniose con una del triaccordo oppure che lo completano in un "tetraaccordo" (ammissibile), ma in entrambi i casi accettando solo le note che appartengono alla "scala" del risolvente,

(iii) le note melodiose con una del triaccordo, ma accettando solo le note che appartengono al triaccordo risolvente,

e per ottenere la "scala" si raccolgono le note (i) (ii) (iii) purchè non siano melodiose con due diverse note del triaccordo.

Infine, per ogni "accordo" le 12 note della tastiera vengono codificate in base alle seguenti proprietà binarie (che la nota può possedere o non possedere):

(1) La nota appartiene al "triaccordo"

(2) La nota completa il "triaccordo" in un "tetraaccordo"

(3) La nota è "armoniosa" con una del "triaccordo"

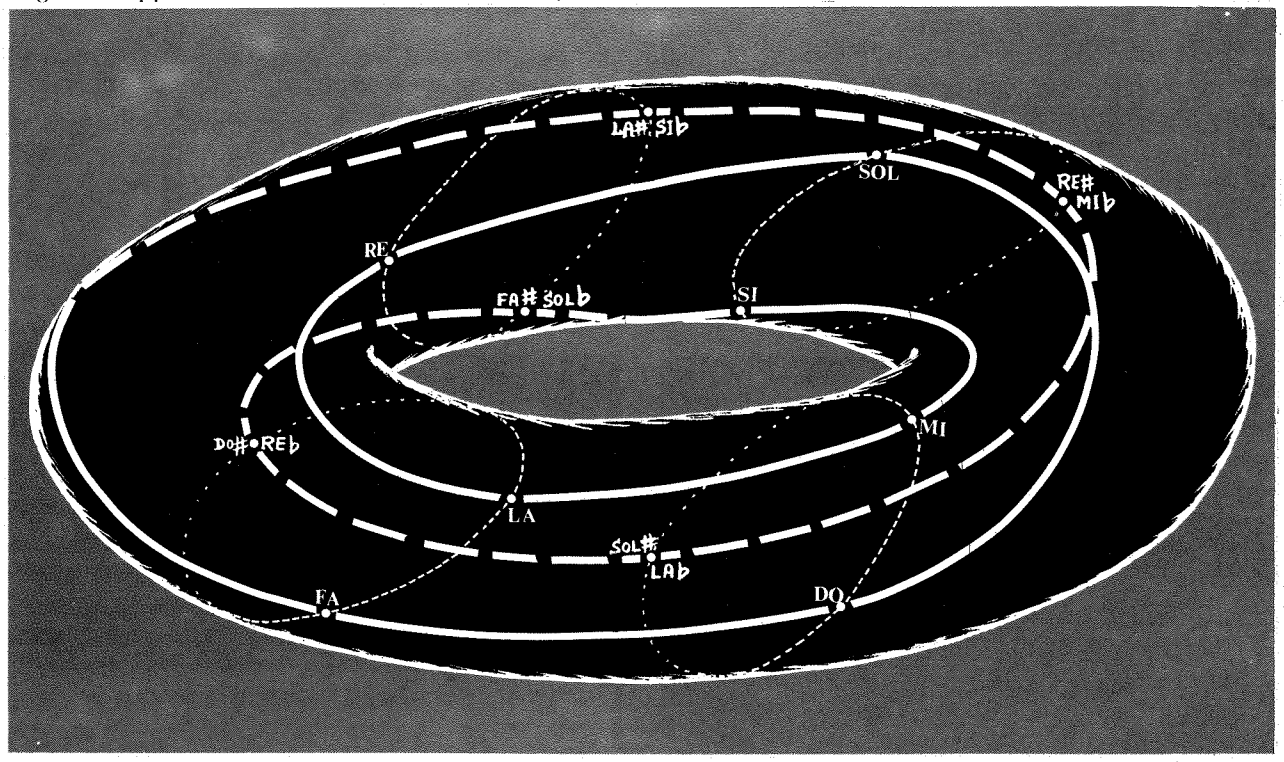
(4) La nota è "melodiosa" con una del "triaccordo"

(5) La nota appartiene alla "scala" dell' "accordo".

Sulla base di queste codificazioni (che permettono di descrivere con un linguaggio matematico: nozioni di armonia, ruolo delle note della melodia rispetto agli accordi dell'accompagnamento, e strutture di brani

(*) Le varie "triadi" vengono qui classificate come "modulazioni" temporanee, e l'accordo "settima di dominante" con omissione di quinta o di terza maggiore come risolvente in tonica maggiore anche se il brano é in minore.

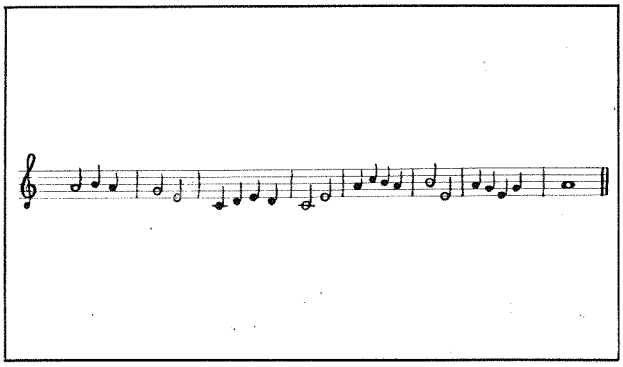
Fig. 4 - Rappresentazione dei legami armonici fondamentali sul toro.



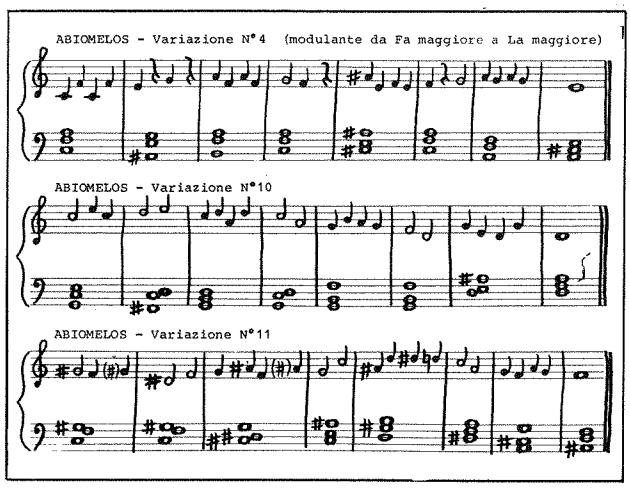
musicali) il calcolatore elettronico ha analizzato numerosi brani ricavati da temi di Beethoven e ne ha estratto matrici di transizioni statistiche che gli permettono di generare nuove musiche seguendo e variando lo schema di "temi" proposti come "input".
Questo algoritmo, chiamato ABIOMELOS, ha composto le musiche eseguite in un concerto per organo, pianoforte e violino il 22.5.1975 presso l'Università dello Stato dell'Oregon, e in una versione modificata ha consentito di effettuare l'esperimento che ora descriviamo.

4 - Esperimento di composizione automatica

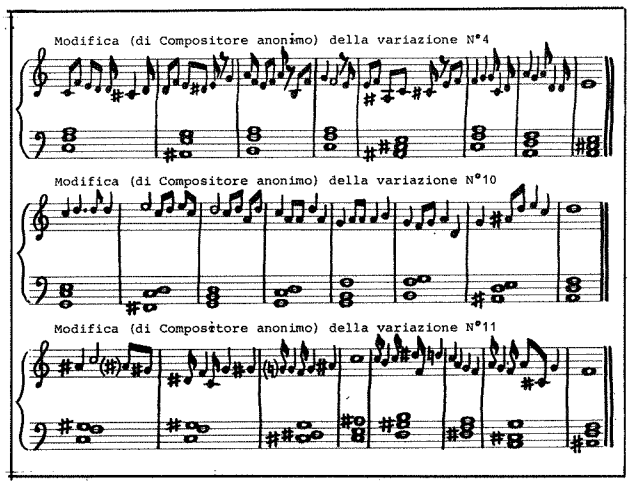
Il 26 aprile 1978 si è tenuto presso il Collegio di S. Caterina a Pavia un esperimento pubblico di composizione automatica di musica di stile classico.
Il pubblico è stato invitato a proporre un "tema". Dei due temi proposti è risultato vincente, in seguito a votazione, il seguente in "la" minore (o "la" dorico?):



Tale tema è stato immediatamente trasmesso alla macchina elettronica (*) dell'Università di Pavia, che in base al programma ABIOMELOS ha composto 14 variazioni, che stampate dalla macchina elettronica sono state subito eseguite sul pianoforte.
La discussione, dopo aver fatto emergere opinioni discordanti anche sul significato di queste ricerche, ha portato a due conclusioni:
(1) Le regole dell'armonia e melodia classica non devono essere "imposte" alla macchina, ma deve essere lasciata una certa "libertà" (casuale) di disubbidire ad alcune, con moderata frequenza (v. le "proibitissime quinte parallele" usate da Mozart e chiamate dai musicisti: "le quinte di Mozart").
(2) Pur raccogliendo in un enorme programma un gran numero di regole desunte dai libri di armonia, da statistiche fatte sui temi di Beethoven, e da elaborazioni fatte dalle macchine elettroniche coi metodi della cosidetta "intelligenza artificiale", è sempre necessaria almeno una selezione umana per rifiutare passaggi musicali che risultano sgradevoli all'uomo.
Una tale selezione, compiuta alla fine della serata, ha portato a scegliere tra le 14 variazioni composte da ABIOMELOS, le tre seguenti (ricopiate usando, come la macchina, in luogo di un b il ♯ della nota precedente):



Successivamente è stato studiato il problema di paragonare il tema proposto inizialmente con la musica che un compositore può scrivere lasciandosi, per così dire, ispirate da queste composizioni automatiche e correggendone alcuni passaggi sgradevoli. Arricchendo la melodia con alcune note tratte quasi sempre dall'accompagnamento della stessa battuta, un compositore che preferisce rimanere anonimo ha con cortese immediatezza scritto le seguenti modifiche rispettivamente delle tre composizioni di ABIOMELOS sopra riportate:



5 - Bibliografia

(1) G. FARINA, "Trattato di armonia", Curci.

(2) E. GAGLIARDO, ABIOSOFOS 1972 - Atti dell'Accademia Ligure di Scienze e Lettere, Vol. XXIX, 1972.

(3) E. GAGLIARDO, "English translation of the paper ABIOSOFOS 1972", Oregon State University Technical Report n. 47, 1973.

(4) L. HILLER, "Form recognition in music", Proc. IFIP Congress - 65, Information processing, 1965.

(5) N. TISCHLER, "Practical Harmony", Allin Bacon, Boston, 1964.

(6) R. ZARIPOV, "Cibernetica e Musica" (in russo), Nauka, 1971; Traduzione italiana: ed. Muzio, Padova, 1979.

(*) Honeywell H6030 che ha impiegato 40 secondi per questa elaborazione.

Evoluzione e prospettive dell'informatica nelle banche italiane

GIOVANNI BOSISIO
Honeywell Information Systems Italia
Servizio Ricerche e Sviluppo Marketing

1 - Introduzione

Il sistema bancario si è andato caratterizzando negli ultimi anni per il progressivo allargamento della sua funzione di intermediazione e per il peso crescente nell'ambito dell'economia.
Risulta infatti che nel 1977 l'espansione del credito totale interno è stata intermediata per oltre l'80% dal sistema bancario, che la raccolta delle banche è aumentata nell'ultimo triennio ad un tasso medio del 23% rispetto al 19% del periodo 70-74, che il rapporto tra depositi e prodotto interno lordo è cresciuto sensibilmente nel periodo più recente, collocandosi al di sopra della linea di tendenza di lungo periodo.
Tale dinamica evolutiva è risultata più accentuata per le banche e casse di risparmio di piccole e medie dimensioni, che hanno così visto aumentare in misura considerevole le loro quote di mercato.
Va peraltro segnalata, in un settore tradizionalmente caratterizzato da una elevata redditività, la presistente contrazione del margine lordo di intermediazione in conseguenza sia di politiche concorrenziali, attuate dalle banche, che hanno inciso sul costo della raccolta, sia di vincoli e massimali che hanno progressivamente frenato la crescita degli impieghi, e determinato un aumento degli investimenti in titoli caratterizzati da livelli di rendimento inferiori.
Anche in termini di redditività risultati migliori sono stati conseguiti dalle banche di dimensioni inferiori.
A fronte di una rapida espansione, le banche si trovano a dover affrontare in prospettiva, problemi di redditività che con ogni probabilità continueranno a manifestarsi a medio termine in quanto strettamente connessi con le condizioni dell'economia italiana ed elementi di maggiore concorrenzialità all'interno del sistema bancario conseguenti sia alla prevista apertura di un numero consistente di nuovi sportelli sia al crescente rilievo assunto dalle medie e piccole banche, ma anche dalle casse rurali e artigiane.
Ne conseguono esigenze di qualificazione del loro ruolo e di razionalizzazione dei modi di operare.
Questi caratteri strutturali ed evolutivi hanno continuato ad incidere sulle tendenze della domanda EDP, come emerge dal presente articolo che si propone di illustrare i risultati di una analisi della evoluzione recente della domanda EDP nel settore bancario e delle tendenze evolutive in atto e previste nel breve e medio termine. (1)

2 - La evoluzione della domanda EDP

La domanda EDP ha continuato a caratterizzarsi nel 1977 per la sostenuta dinamica evolutiva.
Il tasso di crescita del parco sistemi "general purpose" in valore, si è mantenuto attorno al livello medio di lungo periodo (19,8%; 20,1% includendo le casse rurali e artigiane).
La dinamica del parco sistemi è derivata principalmente da ampliamenti e sostituzioni di unità centrali.
I sistemi aggiuntivi (30, in maggioranza piccoli sistemi) hanno invece contribuito solo in misura trascurabile.
Nel corso del 1977 hanno costituito un centro EDP 17 aziende di credito (di cui 10 casse rurali e artigiane).
Risulta evidente che, pur in presenza di livelli di spesa EDP rilevanti, la domanda nel settore è tuttora assai lontana da livelli di saturazione relativa.
La dinamica è però risultata sensibilmente diversificata per fasce di dimensione:

(1) L'analisi, relativa alle aziende di credito ordinario, si basa su informazioni riferite a dicembre '77 raccolte per mezzo di questionari e di interviste dirette e telefoniche
Un questionario postale inviato a tutte le aziende di credito ordinario (escluse le Casse Rurali e Artigiane) dotate di un centro EDP ha avuto un tasso risposta del 25%.
Considerato il tasso di risposta eccessivamente limitato, si è proceduto ad una indagine campionaria sul 50% dell'inverso con interviste telefoniche, salvo per quelle aziende che, incluse nel campione, avevano già fornito informazioni.
Il campione, stratificato sulla base delle dimensioni in tre fasce, è stato costituito con estrazione casuale.
La raccolta di informazioni è stata completata con interviste telefoniche ad un campione di Casse Rurali volta a rilevare gli orientamenti futuri della meccanizzazione EDP.

Tabella 1 - Banche: Evoluzione parco sistemi e terminali 1977 (*)

Mezzi amministrati (miliardi)	Tasso di crescita 1977 %			Tasso medio annuo 1973-1976
	Sistemi	Terminali	Totali	Sistemi
< 100	17.8	87.3	23.5	15.0
100-500	32.4	108.2	42.4	17.3
> 500	16.0	78.5	31.3	21.8
Totale	19.8	82.6	33.3	20.6

(*) Il parco è stato valutato in punti. 1 punto equivale a circa 1 dollaro di canone di locazione mensile.

- Per le banche con oltre 500 miliardi di mezzi amministrati (2) il tasso di crescita (16%) si è mantenuto al di sotto del trend di medio-periodo.

La dinamica è risultata ancor più contenuta (10%) per le 10 maggiori banche.

Pur nell'ambito di realtà differenziate, sembrano emergere in questa fascia sintomi di una certa saturazione.

È proprio in questa fascia che risulta più immediata l'esigenza di qualificare l'utilizzo delle strutture informatiche e di realizzare applicazioni avanzate orientate a supportare le funzioni gestionali e decisionali.

- Le banche medio-piccole (100-500 miliardi) hanno registrato una crescita particolarmente accentuata (32.4%) qualificandosi come il segmento di mercato più dinamico.

Tale andamento è determinato in misura rilevante dalla rapida diffusione del tempo reale di sportello che va assumendo carattere di generalizzazione in questa fascia. L'automazione degli sportelli in tempo reale ha infatti rilevanti impatti in termini di hardware di sistema, oltre naturalmente a riflettersi sulla domanda di terminali. (3)

- Le piccole banche hanno registrato un tasso di crescita più contenuto (17,8%) ma comunque superiore al trend di medio periodo.

Questa fascia è pure caratterizzata attualemente dalla progressiva diffusione del tempo reale di sportello che procede però con maggiore lentezza e con un certo ritardo di adeguamento ai comportamenti delle banche della fascia superiore.

L'aspetto più rilevante dell'evoluzione recente è costituito dalla rapida diffusione dei terminali (+82,6% nel 1977) che ha interessato con connotati sostanzialmente uniformi tutte le banche, determinando un sensibile mutamento del peso relativo dei terminali rispetto ai sistemi (il rapporto, in valore, è passato nell'anno dal 27% ad oltre il 40%, con livelli varianti dal 14% per le piccole banche al 50% per le banche con oltre 500 miliardi di mezzi amministrati).

Come risultato, il tasso di crescita globale della domanda EDP nel settore bancario (sistemi + terminali) è ammontato al 33%. Tale percentuale peraltro sottovaluta il tasso reale di crescita, dato che nella presente analisi non sono stati considerati i minielaboratori il cui utilizzo, sia pure tuttora abbastanza limitato, è però in fase di progressiva espansione.

La diffusione del tempo reale di sportello è proseguita con notevole intensità nel 1977 continuando a rappresentare il principale fattore di espansione della spesa EDP.

Si è inoltre ampliata la gamma di procedure operanti in tempo reale.

Accanto a conti correnti e depositi risultano infatti significative le percentuali relative a anagrafe, estero merci, contabilità di filiale, titoli, fidi, per le quali è peraltro prevista una ulteriore diffusione a breve termine. (Tab. 5)

Le previsioni di ulteriore diffusione e di estensione della rete di sportelli collegati per le banche già dotate dell'applicazione qualificando il tempo reale di sportello come una struttura in fase di progressiva generalizzazione che caratterizzerà l'evoluzione della domanda EDP nel medio termine.

Va sottolineata, per quanto riguarda le modalità di attuazione del tempo reale, la crescente diffusione di terminali intelligenti e di concentratori di linea che determinano una razionalizzazione delle trasmissioni e possono garantire una operatività sia pure limitata in caso di caduta del sistema centrale o della linea di trasmissione, pur non modificando le caratteristiche di accentramento del centro EDP.

Circa 1/3 delle banche operanti in tempo reale utilizza attualmente tali strumenti.

Si va d'altro canto delineando un concreto e crescente interesse per la realizzazione di strutture informatiche decentrate (informatica distribuita) per l'automazione degli sportelli in tempo reale, anche se resta prevalente la tendenza a mantenere l'accentramento.

Su questo tema, si ritornerà più ampiamente in seguito.

3 - Valutazione economica degli investimenti in informatica

È indubbio che la meccanizzazione EDP abbia prodotto effetti positivi sulle attività operative, sull'organizzazione e sulla gestione bancaria.

Limitandoci ad esempio al tempo reale di sportello basti pensare ai benefici derivanti da:

- riduzione della circolazione di documenti cartacei

(2) Per la classificazione si è fatto riferimento ai mezzi amministrati a fine 1976. Le tabelle si riferiscono alle aziende di credito ordinario con esclusione delle Casse Rurali.

(3) cfr. G. Bosisio - L'elaborazione elettronica nel settore bancario in Italia - in Quaderni di Informatica N. 1 - 1977, pag. 43.

Tabella 2 - Settore banche: Dinamica 1977

Mezzi amministrati (Miliardi)	Banche mecc. (*)		Numero sistemi		N. terminali		N. termin. per banca	
	Dic. '76	Dic. '77	Dic. '76	Dic. '77	Dic. '76	Dic. '77	Dic. '76	Dic. '77
100	117	121	128	137	173	319	1.4	2.6
100-500	105	106	137	153	1177	2062	11.2	19.4
> 500	52	52	140	146	7495	12770	144.0	245.0
Totale	274	279	405	435	8845	15151	32.2	54.3

(*) Sono comprese 7 nuovi utenti;
2 banche meccanizzate nel 1976 non compaiono perchè assorbite o sciolte durante il 1977.

Tabella 3 - Settore banche: Banche con tempo reale di sportello - Dinamica evolutiva

Mezzi amministrati (Miliardi)	Set. '76	Dic. '77	Dic. '79	Non Interessate
< 100	6% (7)	22% (27)	56% (67)	44% (54)
100-500	28% (30)	56% (59)	83% (88)	17% (18)
> 500	67% (35)	85% (44)	97% (51)	3% (1)
Totale	27% (72)	47% (130)	75% (206)	25% (73)

Nota: I numeri tra parentesi rappresentano il numero di banche stimato su base campionaria

- riduzione degli errori
- risparmio di personale allo sportello e possibilità di riconversione ad altre attività
- possibilità di svolgere una serie di operazioni di competenza delle filiali direttamente dal centro EDP, (es. assegni di stanza) alleggerendo i carichi di lavoro.

È però un dato di fatto che le decisioni di investimenti in informatica, sia nella fase iniziale, sia nella fase di passaggio al tempo reale e di ulteriori sviluppi, generalmente non sono state supportate da una valutazione economica in termini di analisi costi-benefici, ma si sono basate su valutazioni di convenienza molto semplificate. Risultano inoltre carenti anche analisi a posteriori sull'effettivo risultato dell'utilizzo di strumenti informatici.

Una prima motivazione di tale atteggiamemo, specifica per il settore bancario, è relativa alla assenza di stretti vincoli in termini di costi e alla rilevanza dell'effetto di immagine aziendale e di marketing connesso all'utilizzo di procedure EDP, che hanno progressivamente qualificato l'acquisizione di strumenti informatici come investimento irrinunciabile e comunque vantaggioso.

Una seconda considerazione è relativa alla inconsistenza e arretratezza dei sistemi di contabilità analitica su cui dovrebbe fondarsi una valutazione economica.

La mancanza di analisi di convenienza relative all'informatica risulta in tal senso strettamente connessa ad una generale carenza, all'interno delle banche, di analisi degli andamenti aziendali.

Esistono infine oggettive difficoltà nel definire e quantizzare i benefici derivanti dall'automazione, in relazione alle caratteristiche prevalentemente organizzativo-gestionali di tali benefici.

Solo per la meccanizzazione iniziale o per il passaggio al tempo reale è possibile definire un beneficio in termini di risparmio di risorse. Ma già nel caso del tempo reale questo elemento risulta non prioritario e non facilmente definibile.

Un esempio a tal proposito viene da un grosso istituto di credito che aveva in un primo tempo definito una soglia di volume di attività di filiale al di sotto della quale l'installazione del tempo reale non permetteva risultati significativi in termini di risparmio di personale e di possibilità di riconversione dello stesso ad altre attività.

Tabella 4 - Banche con tempo reale di sportello
Distribuzione secondo la % di sportelli collegati.
Dinamica evolutiva

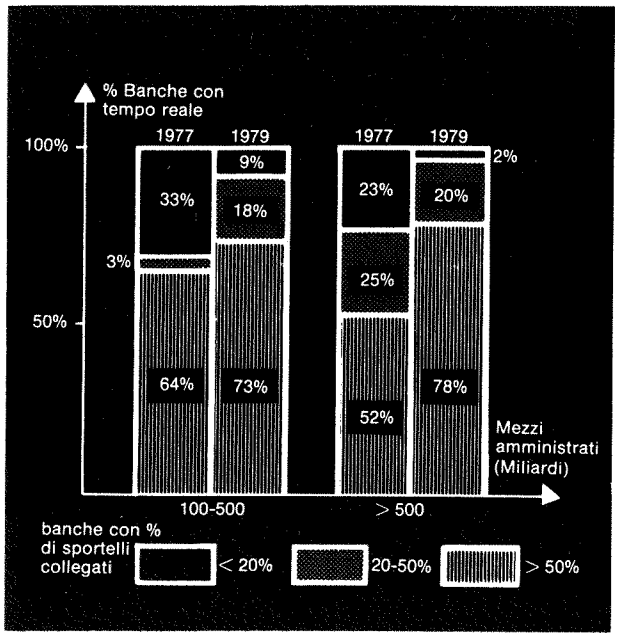
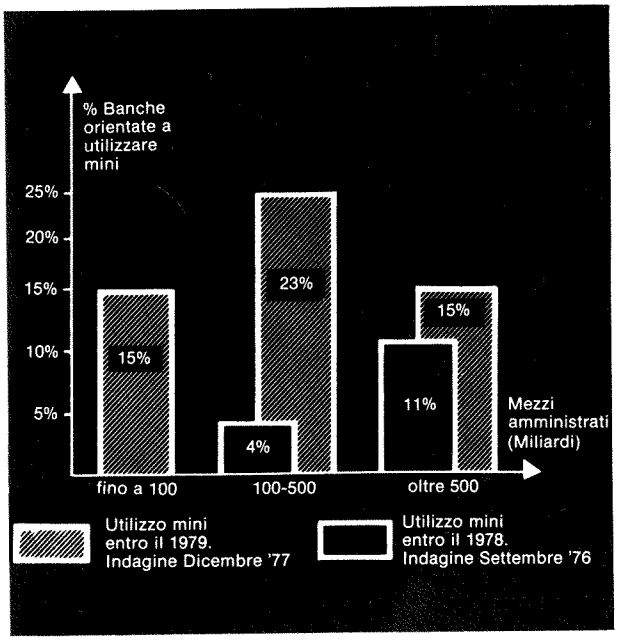


Tabella 5 - Settore banche: Applicazioni in tempo reale - % banche con procedura T.R.

Descrizione	Dicembre 1977			Entro il 1979		
	< 100	100-500	> 500	< 100	100-500	> 500
Anagrafe generale	–	14	48	18	48	76
Conti correnti	22	56	85	56	83	97
Depositi	13	33	45	36	67	68
Gestione fidi	–	10	16	18	43	44
Portafoglio	–	–	4	14	10	8
Estero merci	–	14	36	5	19	60
Contabilità filiali	5	14	20	9	29	32
Contabilità generale	–	5	12	5	14	24
Titoli	–	–	16	–	5	36
N. banche Dic. 1977	121	106	52	121	106	52

In seguito i vantaggi organizzativi, non quantizzabili, derivanti dall'omogeneità delle modalità operative in tutto l'istituto hanno però determinato il superamento di quel criterio e la decisione di estendere la rete di terminali a tutti gli sportelli.
Un tentativo di analisi di prima approssimazione tendente a individuare, per gruppi di banche omogenee, delle connessioni tra livello di meccanizzazione e indici di efficienza operativa e gestionale non ha evidenziato alcuna uniformità di rilevo.
Pur considerando la scarsità dei dati a disposizione e la molteplicità di variabili che agiscono sul livello di efficienza, sembra possibile ipotizzare, data la pratica insussistenza di relazioni, un effetto non univoco dell'utilizzo di strumenti informatici, in relazione al livello qualitativo delle procedure realizzate e al livello di utilizzo delle stesse.
Le condizioni di mercato in cui si trovano ad operare, unitamente agli elevati livelli raggiunti dalla spesa

Tabella 6 - % Banche orientate all'utilizzo di minielaboratori per il tempo reale di sportello entro 2 anni. Confronto tra due indagini.



EDP, pongono peraltro alle banche problemi non oltre dilazionabili di razionalizzazione delle modalità operative e gestionali, di razionale utilizzo delle stesse strutture informatiche e quindi di valutazione economica degli effetti di investimenti in informatica.
Condizioni essenziali per realizzare questi obiettivi sono, innanzitutto, uno sforzo di standardizzazione delle procedure che consenta di qualificare e rendere più facilmente trasferibili le applicazioni e le esperienze relative e la realizzazione di sistemi di contabilità analitica dei costi e ricavi.
Nell'attuale situazione entrambi questi problemi sono ben lontani dall'essere affrontati con impegno e risolti. Esiste una diffusa convinzione della necessità di affrontarli che tarda però a tradursi in azioni concrete.

4 - Tendenze evolutive

Nel medio termine continuerà verosimilmente a manifestarsi una accentuata dinamica evolutiva in relazione principalmente a:

- consolidamento delle strutture EDP esistenti e adeguamento agli accresciuti volumi di attività. Questo processo potrà risultare particolarmente vivace per le banche medie e piccole sia per lo stato di fluidità della meccanizzazione che per la tendenza a espandersi che le caratterizza.
- generalizzazione del tempo reale di sportello
- riduzione del gap temporale tra introduzione di nuove applicazioni e loro diffusione.

È stato sottolineato come spesso la diffusione di procedure informatiche sia frutto di un processo imitativo che, innescato dalle banche maggiori, coinvolge progressivamente, con un certo ritardo, le altre banche.
L'osservazione degli andamenti più recenti dimostra che questi gap temporali, in relazione al crescente livello di maturità della meccanizzazione, si vanno sensibilmente riducendo.
Ne deriva la possibilità di una diffusione più rapida di applicazioni e realizzazioni future, rispetto alle esperienze attuali e passate.
L'indagine ha permesso inoltre di rilevare alcune specifiche linee di tendenza lungo le quali è prevedibile che

Tabella 7 - % Banche orientate a utilizzare minielaboratori entro fine 1979

Mezzi amministrati (Miliardi)	Mini per tempo reale decentrato	Mini dedicati
100	15% (19)	8% (9)
100-500	23% (24)	13% (14)
> 500	15% (8)	39% (20)
Totale	18% (51)	15% (43)

Nota: I numeri tra parentesi rappresentano il numero di banche stimato su base campionaria

Tabella 8 - Meccanizzazione del Servizio Estero

Mezzi amministrati (Dicembre '76) (Miliardi)	Banche Agenti	Banche con Servizio Estero		
		Non meccanizzato	Mecc. con elab. centrale	Mecc. con mini
50-100	45	58% (26)	8% (4)	33% (15)
100-250	59	42% (25)	33% (19)	25% (15)
250-500	35	33% (12)	33% (12)	33% (12)
> 500	46	-	90% (41)	10% (5)
Totale	185	34% (63)	41% (76)	25% (47)

Nota: I numeri tra parentesi rappresentano il numero di banche stimato du base campionaria

possa indirizzarsi lo sviluppo dell'informatica nelle banche:
1) Decentramento delle strutture informatiche (informatica distribuita) sia per quanto riguarda l'automazione degli sportelli in tempo reale sia per la meccanizzazione si specifiche procedure.
2) Introduzione e ampliamento di sistemi elettronici di trasferimento di fondi tendenti a ridurre la circolazione di documenti cartacei di pagamento.
3) Applicazioni gestionali - contabilità analitica, ricerca operativa, marketing.
4) Utilizzo crescente di prodotti applicativi standard in relazione soprattutto all'espansione e alla qualificazione della meccanizzazione nelle medie e piccole banche e all'affermarsi dell'informatica distribuita.

• **L'informatica distribuita** intesa come dislocazione di impianti di elaborazione presso i singoli sportelli, incontra un interesse crescente presso le banche, anche se le installazioni sono tuttora contenute.
Il confronto tra gli orientamenti attuali e quelli espressi in occasione di una precedente indagine, evidenzia questa rilevante crescita di interesse che sembra riguardare tutte le fasce, con una particolare accentuazione per le banche medie e piccole. (Tab. 6)
Si può stimare, sulla base delle informazioni rilevate, che circa il 20% delle banche operanti in tempo reale avrà entro il 1979 una struttura del centro EDP decentrata.
Il più vasto interesse in tal senso da parte di banche medie e piccole deriva dallo stato di fluidità dell'automazione in tempo reale degli sportelli in questa fascia.
Il maggior consolidamento dell'applicazione determina invece più accentuate resistenze al cambiamento nelle banche medio-grandi.

Va inoltre sottolineato che la lentezza della domanda nell'orientarsi verso l'informatica distribuita è stata determinata in misura considerevole dallo stesso atteggiamento delle principali case fornitrici che solo recentemente si sono orientate a predisporre offerte specifiche per questa area applicativa.
L'utilizzo di minielaboratori rappresenta, a differenza di strumenti quali concentrati di linea e terminali intelligenti, una sostanziale modifica delle strutture informatiche in quanto trasferisce la potenza di calcolo e gli archivi presso i singoli punti operativi, pur consentendo il mantenimento di un controllo centralizzato della rete.
La soluzione distribuita è avvalorata da una serie di considerazioni tecniche, organizzative ed economiche:
• incremento delle prestazioni e maggiore disponibilità complessiva del sistema in conseguenza delle ridotte possibilità di rottura simultanea di tanti centri fisicamente indipendenti.
Tale caratteristica assume particolare rilievo per il tempo reale di sportello garantendo una maggiore sicurezza. L'interconnessione tra le varie unità garantisce poi la continuità operativa evitando l'interruzione dell'attività in caso di caduta di uno dei sistemi.
• possibilità di adeguamento del sistema e delle modalità operative alle esigenze delle diverse sedi, e quindi di attuazione di una parziale autonomia di filiale.
• riduzione dei costi complessivi (hardware e trasmissioni) consentita da un certo livello ottimale di distribuzione.
La realizzazione di una struttura decentrata per il tempo reale di sportello può rappresentare una soluzione ottimale oltre che per singole banche, per consorzi di piccole banche.
Una struttura basata su un elaboratore centrale con-

sortile e una serie di mini decentrati può consentire infatti ad ogni singola banca di operare in condizioni di autonomia pur continuando a sfruttare le economie di scala connesse alla elaborazione accentrata.
Un discorso a parte va fatto per le Casse Rurali (660 in totale) concentrate principalmente in alcune aree geografiche, caratterizzatesi nel periodo più recente per la rapida espansione del volume di mezzi amministrati e la notevole qualificazione del loro ruolo.
La forma di meccanizzazione EDP prevalente è rappresentata dai centri consortili provinciali e regionali promossi dall'Istituto Centrale di Categoria (ICCREA) che consentono indubbie economie di scala.
L'indagine condotta evidenzia che:

- gli strumenti e le modalità di collegamento fra ogni singola cassa e il centro consortile pongono in molti casi dei limiti alla tempestività delle informazioni
- le esigenze di autonomia e di elasticità gestionale avvertite con intensità dalle singole casse sono spesso contrastate dalla centralizzazione delle strutture informatiche.

Con riferimento a questi problemi emergono due atteggiamenti divergenti, motivati generalmente dal livello di servizio garantito dal centro consortile:

- da un lato, viene espressa l'esigenza di strumenti informatici decentrati (minielaboratori) che garantiscano una larga sfera di autonomia operativa e gestionale, pur nell'ambito del centro consortile.
- dall'altro lato vengono espresse spinte centrifughe in direzione di centri autonomi propri o di piccoli centri consortili.

Risulta inoltre rilevante l'orientamento a utilizzare minielaboratori dedicati per specifiche procedure - estero merci, titoli -, proprio in relazione alla rilevante autonomia organizzativa e gestionale che caratterizza tali servizi. (Tab. 7)
L'utilizzo di mini dedicati riduce inoltre in questo caso l'impatto dell'applicazione sull'elaboratore centrale e sulle procedure esistenti semplificandone la realizzazione.
Nel caso specifico della procedura estero merci risultano già ampiamente diffusi questi sistemi, acquistati generalmente insieme al relativo software applicativo. Tale tendenza appare in fase di rafforzamento. (Tab. 8)
L'impiego di minielaboratori si prospetta poi assai ampio in relazione all'entrata in funzione e alla prevista espansione della rete SWIFT e alla realizzazione di altri strumenti si trasferimento elettronico di fondi.

- **Le applicazioni di trasferimento elettronico di fondi (EFTS)** diffuse in misura assai limitata presso le banche italiane potranno avere un ruolo di rilievo nei prossimi anni.

Con questa espressione si fa riferimento a una serie di soluzioni volte a ridurre la circolazione di documenti cartacei per le operazioni di pagamento:

- automazione dei pagamenti (riscossioni) periodici di massa
- stanze automatiche di compensazione interbancaria
- automazione dei regolamenti internazionali
- sportelli bancari automatici
- automazione dei pagamenti presso i punti di vendita.

Di rilievo risultano le realizzazioni e gli sviluppi in atto all'estero in questa area.
In Giappone e negli Stati Uniti, sono funzionanti reti estese per i trasferimenti interbancari e stanze di compensazione automatiche e sono ampiamente diffusi sportelli bancari automatici e terminali presso i punti di vendita.
Tali strumenti sono in fase di espansione verso una rete integrata di pagamenti automatici riguardante banche, imprese, enti pubblici, punti di vendita, consumatori.
Anche in Europa sono attualmente in funzione analoghi strumenti, sia pure con un livello di diffusione inferiore.
La situazione in Italia è al momento attuale, sensibilmente più arretrata.
Oltre a sistemi automatici per pagamenti periodici sono utilizzati, in misura tuttora limitata, i cash dispenser, prevalentemente da parte di grandi banche.
L'esperienza più significativa è costituita dall'adesione di numerose banche al sistema SWIFT per i trasferimenti internazionali, funzionante dal 1977.
L'interesse per lo SWIFT, che consente un collegamento diretto con numerose banche straniere determinando una qualificazione del servizio, si è rapidamente esteso, tanto che è prevedibile un rilevante incremento di adesioni a breve termine.
Emerge inoltre una consistente propensione a utilizzare in futuro cash dispenser. Il 20% delle banche (oltre il 50% delle banche oltre 500 miliardi di mezzi amministrati) intendono utilizzare questi strumenti a medio termine.
Non è invece considerata attuabile in tempi brevi l'installazione di impianti di validazione delle carte di credito e di pagamento automatico presso i punti di vendite.
La realizzazione di una rete per regolamenti interbancari interni rimane legata alla attuazione dei progetti SIA e STACRI.

- Per quanto riguarda la **contabilità analitica** circa il 15% delle banche italiane (oltre un terzo delle banche con oltre 500 miliardi di mezzi amministrati) esprimono interesse e affermano di avere in progetto realizzazioni di questo tipo.

Al momento attuale le realizzazioni concrete risultano peraltro assai limitate e parziali.

In conclusione si può prevedere a medio termine il proseguimento di un tasso di crescita attorno al 20% per i sistemi general purpose e una rapida espansione dei terminali e dei minielaboratori.
Lo sviluppo potrà essere ulteriormente accentuato da una qualificata offerta di prodotti applicativi.

Collana dei Quaderni di informatica

Novità

D. Ferrari / G. Serazzi / A. Zeigner
Le prestazioni degli elaboratori elettronici
Misura/Valutazione/Ottimizzazione

Uno degli aspetti di maggior rilievo nei più recenti sviluppi dell'informatica è il crescente approfondimento degli studi sulle prestazioni dei sistemi informatici alla cui straordinaria crescita si è via via affiancata la diffusione sempre più ampia di tecniche, metodi e strumenti atti a indagarne gli aspetti particolari.
Rispetto ad altri sull'argomento, questo volume presenta due caratteristiche salienti: nella sua impostazione di fondo persegue l'obiettivo di dare al lettore tutti gli elementi fondamentali riguardanti i metodi e gli strumenti per operare con efficacia nel campo della valutazione dei sistemi informatici, mentre nello sviluppo della materia dà largo spazio agli aspetti pratici della misurazione di attività dell'hardware e del software degli impianti di elaborazione dei dati. Non vengono inoltre tralasciati i problemi economici riguardanti la redditività degli investimenti nelle operazioni di *tuning* e di controllo delle prestazioni.
Il libro colma inoltre la sentita lacuna della mancanza di testi in italiano sull'argomento sviluppando in maniera organica quanto è possibile trovare sparso in numerose pubblicazioni estere.
La struttura del volume consente due chiavi di lettura: quella informativa di carattere generale, particolarmente adatta a chi vuole accostarsi all'argomento, e quella più approfondita per chi desidera integrare le proprie conoscenze della materia.

Sommario

Definizione del problema: Definizioni e concetti di base - Gli obiettivi della valutazione - I sistemi di riferimento - Gli indici di prestazione (il carico (workload); il tempo di turnaround; il tempo di risposta; la produttività) - Tecniche di valutazione.

Tecniche di misurazione: Misurazioni per rilevamento di evidenti - Misurazioni con carichi pilota (benchmarking) e loro costruzione - Misurazioni per campionamento - Determinazione del campione e accuratezza della stima - Simulazione.

Rappresentazione delle grandezze misurate: Tabelle e grafici - Profili di utilizzo (o di Gantt) - Grafici di Kiviat - Forme standard dei grafici di Kiviat.

Strumenti di misurazione: Le caratteristiche principali di uno strumento - Strumenti Sofware (software/monitors) - Un esempio di campionamento (sampler) - Misure di tempo, orologi e temporizzatori - Strumenti hardware (hardware monitors) - Esempi di applicazioni di strumenti hardware - Simulatori.

Metodologie di valutazione: Definizione degli obiettivi - La scelta dello strumento - La pianificazione del lavoro - Ricerca delle strozzature (bottlenecks) - L'eliminazione delle strozzature.

Ottimizzazione dei sistemi: Il bilanciamento di un sistema multiprogrammato (caso 1: strozzature nei dischi e nei canali; caso 2: insufficiente capacità di memoria; caso 3: eccesso di capacità in un impianto con due sistemi) - L'ottimizzazione di un sistema interattivo (caso 1: sovraccarico dei canali; caso 2: inefficiente ripartizione del carico tra due sistemi; caso 3: ottimizzazione della disposizione degli archivi su disco) - Il miglioramento di un sistema con memoria virtuale (caso 1: riduzione della paginazione in un impianto con due sistemi a memoria virtuale) - Controllo delle prestazioni di un sistema per la gestione di banche di dati - Bilanciamento ottimale del carico su unità di I/O con un modello analitico.

Ottimizzazione del carico: Criteri generali e scelta dei programmi da ottimizzare - Tipi di ottimizzazioni dei programmi - Gli indici di prestazione del carico - Riduzione del tempo di esecuzione (miglioramento del codice; miglioramento dell'attività di I/O; riduzione della frequenza di paginazione).

Considerazioni economiche: Costi e ricavi di uno studio di valutazione (caso 1: riduzione della configurazione hardware; caso 2: rinvio dell'installazione di un nuovo sistema; caso 3: ottimizzazione di un programma di sistema; caso 4: ottimizzazione dei programmi degli utenti; caso 5: scelta di uno strumento per il controllo dell'attività di banche di dati).

Appendice A: Nozioni di fisica del software (K.W. Kolence): Il punto di vista della fisica del software - Sistemi di base (unità di software; configurazioni, sottoconfigurazioni ed elaboratori equivalenti; le classi di componenti) - Proprietà fondamentali (il lavoro software: il tempo; lo spazio di memoria occupato) - La potenza software.

Appendice B - Sinonimi e glossario